Characterization of mesophyll-specific promoters for C_4 engineering of rice and mutational analysis of leaf anatomy in *Arabidopsis thaliana*

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

Presented by

Shipan Das Gupta

from Chittagong, Bangladesh

Düsseldorf, September 2018

from the institute of “Developmental and Molecular Biology of Plants”
at the Heinrich Heine University Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: **Prof. Dr. Peter Westhoff**
Co-supervisor: **Prof. Dr. Maria von Kroff Schmising**

Date of the oral examination: 17th December, 2018

Affirmation for the Doctoral Thesis

I herewith declare that I have written this thesis independently and myself. I have used no other sources than those listed. I have indicated all places where the exact words or analogous text were taken from sources. I assure that this thesis has not been submitted for examination elsewhere.

Düsseldorf, 14th September 2018

Shipan Das Gupta

Table of contents

Abbreviation

A	Adenine
A. thaliana	*Arabidopsis thaliana*
ATP	Adenosine triphosphate
bp	Base pairs
BS	Bundle sheath
°C	Degree Celsius
C_1, C_2, C_3, C_4	One-, two-, three-, four-carbon molecule
C_4Ppc	C_4-type phospho*enol*pyruvate carboxylase gene
C	Cytosine
CA	Carbonic anhydrase
CCM	CO_2-concentrating mechanism
CDS	Coding DNA sequence
CO_2	Carbon dioxide
DIT1+2	Dicarboxylate transporter 1+2
D. sanguinalis (Ds)	*Digitaria sanguinalis*
DsPpc	Phospho*enol*pyruvate carboxylase gene of *D. sanguinalis*
DNA	Deoxyribonucleic acid
EMS	Ethyl methanesulfonate
F. trinervia (Ft)	*Flaveria trinervia*
G	Guanine
GDC	Glycine decarboxylase complex
GFP	Green fluorescent protein
GGT	Glutamate-glyoxylate aminotransferase
$GLDPA_{Ft}$	Glycine decarboxylase PA gene of *Flaveria trinervia*
$GLDT_{Ft}$	Glycine decarboxylase T gene of *Flaveria trinervia*
GLYK	Glycerate kinase
GOX	Glycolate oxidase
GUS	β-Glucuronidase
h	Hour(s)
HCO_3^-	Bicarbonate (hydrogen carbonate)
HPR	Hydroxypyruvate reductase
Kb	Kilobases

kDA	Kilo Dalton
LB	T-DNA left border
MDH	Malate dehydrogenase
M	Mesophyll
M1, M2, M3, M4	1st, 2nd, 3rd or 4th mutant plant generation
ME	Malic enzyme
min	Minute(s)
mRNA	Messenger ribonucleic acid
MU	4-Methylumbelliferone
MUG	Methylumbelliferyl-β-D-galactopyranoside
N_2	Molecular nitrogen
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NH_3	Ammonia
nt	Nucleotides
O_2	Molecular oxygen
OAA	Oxaloacetate
ORF	Open reading frame
OsO_4	Osmium tetroxide
PCR	Polymerase chain reaction
PEP	Phospho*enol*pyruvate
PEPC	Phospho*enol*pyruvate carboxylase
P. miliaceum (Pm)	*Panicum miliaceum*
PmPpc	Phospho*enol*pyruvate carboxylase gene of *P. miliaceum*
*ppc*A	Phospho*enol*pyruvate carboxylase gene A of *Flaveria*
3-PGA	3-Phosphoglycerate
2-PG	2-Phosphoglycolate
PGLP	Phosphoglycolate phosphatase
PPDK	Pyruvate, orthophosphate dikinase
RNA	Ribonucleic acid
Rubisco	Ribulose 1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-bisphosphate
s	Second(s)
S. italica (*Si*)	*Setaria italica*

SiPpc	Phospho*enol*pyruvate carboxylase gene of *S. italica*
SGT	Serine:glyoxylate aminotransferase
SHMT	Hydroxymethyltransferase
SNP	Single-nucleotide polymorphism
S. viridis (*Sv*)	*Setaria viridis*
SvPpc	Phospho*enol*pyruvate carboxylase gene of *S. viridis*
T	Thymine
T-DNA	Transfer DNA
TEM	Transmission electron microscopy
TF	Transcription factor
THF	Tetrahydrofolate
U. maxima (*Um*)	*Urochloa maxima*
UmPpc	Phospho*enol*pyruvate carboxylase gene of *U. maxima*
X-Gluc	5-bromo-4-chlor-3-Indoyl-β-D Glucuronid
Z. mays (*Zm*)	*Zea mays*
ZmPpc	Phospho*enol*pyruvate carboxylase gene of *Z. mays*

I. Introduction

1. A general introduction to C_3 Photosynthesis and Photorespiration reaction

1.1. Role of the bi-functional enzyme Rubisco in Photosynthesis

Photosynthesis is a complex biochemical mechanism performed by plants, algae, and some bacteria. Plants can perform photosynthesis using chloroplasts, a specialized cell organelle that converts atmospheric CO_2 and solar energy into chemical energy. Life on earth is directly or indirectly dependent on plants' photosynthesis as it is the ultimate source of all humankind's food and oxygen. The complex process of photosynthesis can be conceptually divided into two-step reactions. The first step is known as 'light-dependent' reaction and the second one is 'light-independent' reaction. In the light-dependent reaction, plant splits the water molecules using the energy of light and produces the molecular oxygen and free electrons. The photolysis of water occurs in the water-splitting complex of photosystem II (PSII), which is embedded in the thylakoid membranes of the chloroplasts. The released molecular oxygen from the photolysis reaction provides the major source of O_2 for the animal kingdom. Later the electrons transfer between the two photosystems (PSII and PSI), which ultimately leads to the reduction of $NADP^+$ and generation of ATP. The produced NADPH and ATP from the light dependent reaction are known as assimilatory energy. The two assimilatory molecules are needed in the 'light-independent' reaction to assimilate atmospheric CO_2 into carbohydrates. The assimilation reaction took place in a cycle of reactions known as Calvin-Benson cycle (Bassham et al., 1950).

Three different mechanisms of photosynthesis can be observed in the plant world. The most common and primitive one is C_3 photosynthesis, from which the other two, C_4 and Crassulacean Acid Metabolism (CAM) cycle have evolved. The differences between the three photosynthetic plants are in their location and CO_2 fixation mechanism. Around 85 % of all land plants, including many agronomically important crop species, follow the C_3 photosynthetic pathway (Ehleringer et al., 1991). Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco, EC.4.1.1.39), one of the key proteins involved in the C_3 photosynthetic pathway is the most abundant protein in the

world (Ellis, 1979). Rubisco comprises approximately 50 % of the total soluble proteins of the C_3 photosynthetic plants (Parry et al., 2003). In higher plants, Rubisco is a hexadecamer enzyme composed of eight nucleus-encoded small subunits (RbcS) and eight chloroplast-encoded large subunits (RbcL) (Andersson et al., 2003; Liu et al., 2010). Four of the RbcL subunits form a tetramer, and two of this type of tetramer arrange in an antiparallel orientation to build the $RbcL_8$ core complex. The $RbcL_8$ core complex is capped by four RbcS subunits at the top and the bottom respectively, forming the cylindrical shape of this enzyme (Newman et al., 1993). During photosynthesis, Rubisco incorporates CO_2 into a five-carbon compound Ribulose 1,5-bisphosphate and produces two molecules of 3-phosphoglycerate (3-PGA) (Calvin & Benson, 1948). The resultant 3-PGA molecules are fed into the Calvin-Benson cycle where the primary components of carbohydrate are synthesized and thus lead to the growth of the plant (Figure 1).

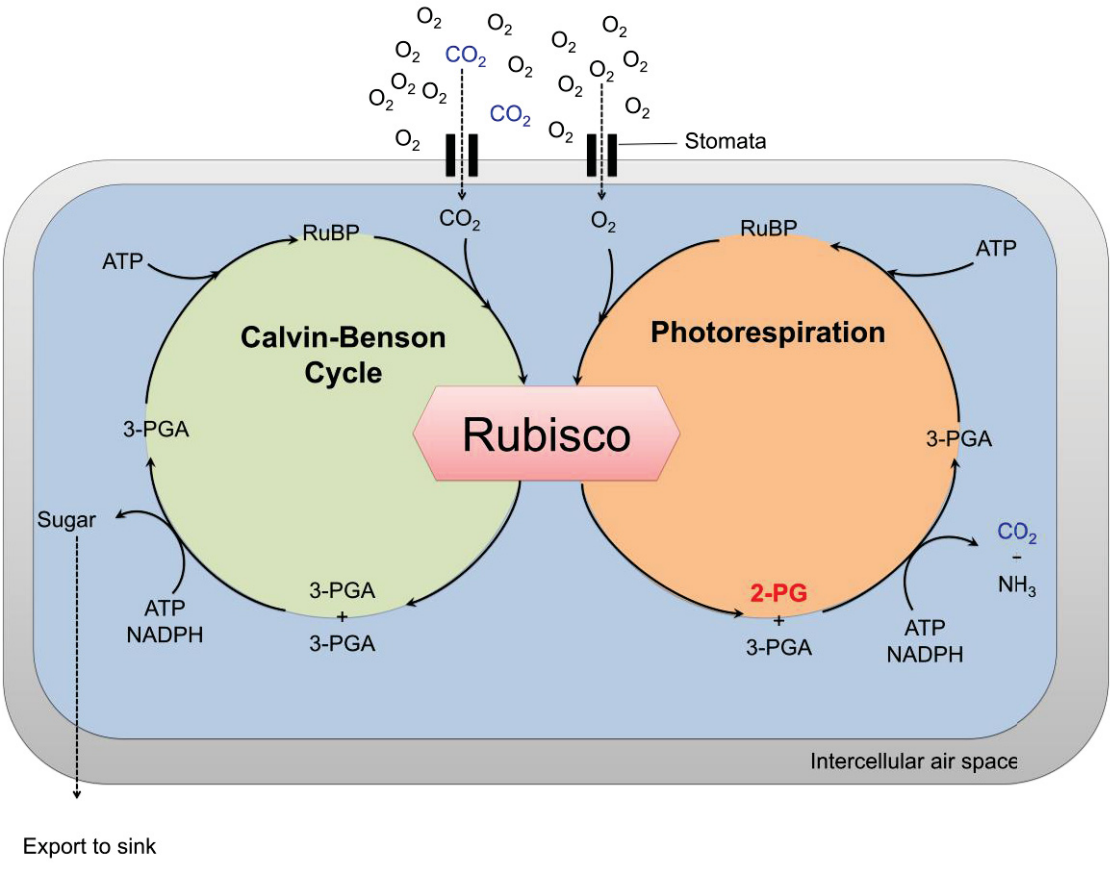


Figure 1: A comparison between the Calvin-Benson cycle and the photorespiration reaction in the C_3 plants. During Calvin cycle Rubisco incorporates CO_2 into Ribulose 1,5-bis phosphate (RuBP) and produces two molecules of 3-phosphoglycerate (3-PGA). In the subsequent steps 3-PGA molecules are used for

synthesis of sugar molecule and for regeneration of RuBP as well. Conversely, in photorespiration O_2 is bound to RuBP and forms 2-phosphoglycolate (2-PG) and 3-PGA, PG then undergoes a series of energy requiring reactions by releasing CO_2 and NH_3.

However, Rubisco is a bi-functional enzyme, it can also add O_2 to Ribulose 1,5-bis phosphate resulting in one molecule each of 3-phosphoglycerate (3-PGA) and 2-phosphoglycolate (2-PG) (Hatch, 1987). 2-PG and its derivatives glyoxylate and glycerate are metabolically toxic for the plant cells (Bowes et al., 1971; Jordan & Orgen, 1984). They are known as inhibitors of many enzymes in the Calvin-Benson cycle and central carbon pathways (Kelly & Latzko, 1976; Campbell & Orgen, 1990; Norman & Colman, 1991). Therefore, the 2-PG must be detoxified immediately from the cells. Recycling of the 2-PG is mediated through a series of energy-consuming reactions called photorespiration. Photorespiration is a complex biochemical reaction, which is distributed between the chloroplasts, mitochondria, peroxisomes, and the cytosol (Bauwe et al., 2010). During photorespiration, two molecules of 2-PG are converted into one molecule of 3-PGA via nine-enzymatic steps reaction at the expense of ATP and NADPH. This pathway is, therefore, also known as photorespiratory C2 cycle. Photorespiration reaction not only demands energy, but also causes net loss of prefix CO_2. During the process of photorespiration in C_3 plants, recycling of the 2-PG to 3-PGA leads to a net loss of 25 % - 30 % CO_2, which are already fixed. In C_3 photosynthetic plants, there is a constant competition between the Calvin-Benson cycle (photosynthesis) and the photorespiration cycle because Rubisco is involved in both the pathways (Figure 1). Depending on the CO_2 availability, Rubisco can act either as carboxylase or oxygenase. When CO_2 is plentiful, Rubisco acts as carboxylase, and the Calvin-Benson cycle converts CO_2 and water into carbohydrate and O_2 molecules. On the other hand, when the level of CO_2 is low, Rubisco acts as oxygenase, activating the photorespiration reaction that consumes energy and releases CO_2. Additionally, the carboxylase/oxygenase activity of Rubisco is greatly influenced by the temperature. The carboxylase activity of Rubisco decreases at temperature above 25º C (Ku & Edwards, 1977a; Jordan & Orgen, 1984; Brooks & Farquhar, 1985). At higher temperatures, the solubility of CO_2 decreases much faster than that of O_2. Therefore, the oxygenase activity of Rubisco is esteemed under unfavorable conditions, including high temperature and dryness, and reduces the photosynthetic efficiency of C_3 plants

by up to 40 % (Ehlerginer & Monson, 1993; Raines, 2011; Fernie et al., 2013). A regional scale model reveals that photorespiration caused reduction of the US soybean and wheat yields by 36 % and 20 % respectively. Also, decreases of only 5 % in the production of yield due to photorespiration would cause losses of approximately $500 million annually in the United States (Walker et al., 2016).

1.2. Photorespiration is a costly but indispensable pathway for performing oxygenic photosynthesis

Photorespiration reaction is an essential pathway for the plant, which makes the best of a bad situation caused by the oxygenase activity of Rubisco. No other pathway which can metabolize the 2-PG exists in plants except the photorespiration pathway. Furthermore, it can rescue the ¾th of carbon in 2-PG. More than 20 different enzymes and transporters (mostly unidentified) spanning in the chloroplast, peroxisome, mitochondria, and cytosol are involved in this complex pathway (Bauwe, 2010; Hagemann & Bauwe, 2016) (Figure 2). The first step of the pathway starts in the chloroplast with de-phosphorylation of a 2-PG into glycolate catalyzes by the 2-PG phosphatase (PGLP, EC 3.1.3.18). Glycolate is then exported from the chloroplast to cytosol via a glycolate/glyoxylate transporter, subsequently diffusing to the peroxisome (Pick et al., 2013). The glycolate oxidase (GOX, EC 1.1.3.15) enzyme in peroxisome performs an irreversible oxidation to glycolate, resulting in a production of glyoxylate. In this reaction, H_2O_2 is produced as a byproduct.

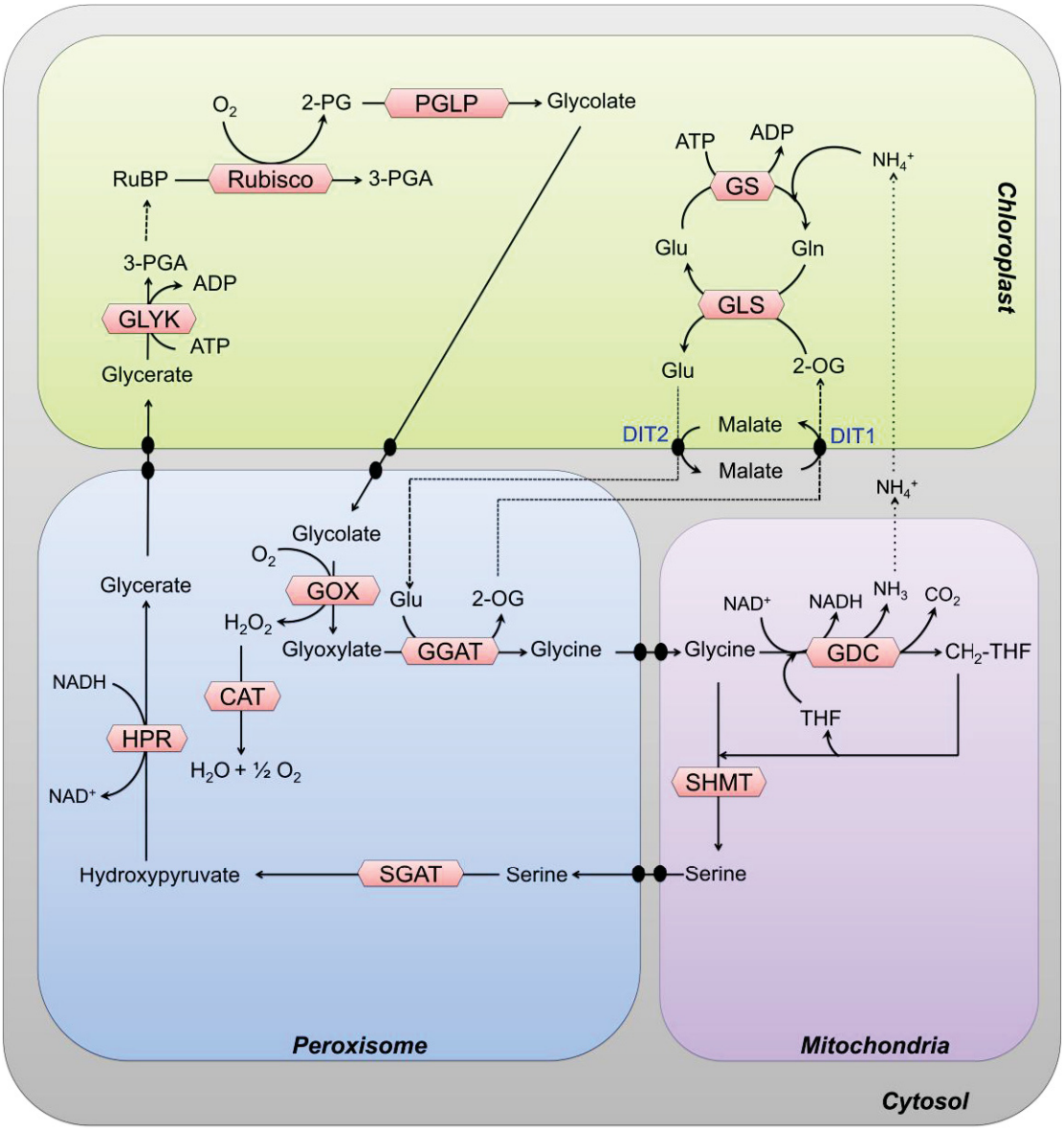


Figure 2: The photorespiration pathway. The photorespiration pathway in higher plants is distributed in the chloroplast, peroxisome, and mitochondria. Enzymes involved in this pathway are represented in hexagonal structure. Here, PGLP, Phosphoglycolate phosphatase; GOX, Glycolate oxidase; CAT, Catalase; GGAT, Glutamate:Glyoxylate aminotransferase; GDC, Glycine decarboxylase complex; SHMT, Serine hydroxymethyltransferase; SGAT, Serine:Glyoxylate aminotransferase; HPR, Hydroxypyruvate reductase; GLYK, D-Glycerate-3-kinase; GLS, Glumate synthase; GS, Glutamine synthetase; DIT-1, Decarboxylate transporter-1; DIT-2; Decarboxylate transporter-2; Gln, Glutamine; Glu, Glutamate; 2-OG, 2-Oxoglutarate; THF, Tetrahydrofolate; PGA, 3-Phosphoglyceric acid. Figure adapted from (Bauwe, 2010; Peterhansel et al., 2010)

In order to prevent the accumulation of reactive oxygen species inside the cells, the

H_2O_2 is immediately degraded into H_2O and O_2 by a catalase (CAT, EC 1.11.1.6). The next step of the photorespiration pathway is the transamination of the glyoxylate to glycine by glutamate:glyoxylate aminotransferase (GGAT, EC 2.6.1.4). Glycine is then transported to mitochondria where it gets decarboxylated and deaminated by the multi-enzyme complex glycine decarboxylase (GDC, EC 1.4.4.2) producing CO_2, NH_3, NADH and 5,10 methylene terahydrofolate (CH_2-THF). The CH_2-THF is used by serine hydroxymethyltranferase (SHMT, EC 2.1.2.1) as the substrate for the transfer of the C1 moiety to another molecule of glycine and resulting in the production of serine. Serine is then transported to the peroxisome and gets deaminated to hydroxypyruvate by serine:glyoxalate aminotransferase (SGAT, EC 2.6.1.45). Hydroxypyruvate is subsequently reduced to glycerate catalyzed by hydroxypyruvate reductase (HPR, EC 1.1.1.29). The final steps of photorespiration reaction take place in the chloroplast where the D-glycerate 3-kinase (GLYK, EC 2.7.1.31) catalyzes the regeneration of 3-PGA from glycerate using the energy of ATP.

Photorespiration allows photosynthesis in an oxygen-containing environment and is, therefore, essential for all oxygenic photosynthetic organisms (Eisenhut et al., 2008; Bauwe, 2010; Rademacher et al., 2016). Although, the photorespiration pathway is energy consuming, it has some positive aspects too. On knocking out different genes of the photorespiratory pathway provides enough evidence that this is an essential process for a plant to grow in an ambient atmosphere condition (Sharkey, 1988; Somerville, 2001; Engel et al., 2007; Bauwe et al., 2012). The photorespiration pathway not only removes the toxic metabolites from the cell but is also involved in providing its intermediate to the synthesis of several amino acids (Novitskaya et al., 2002). It has been reported in many studies that there is a positive correlation between the photorespiration and nitrogen uptake in wheat and Arabidopsis (Rachmilevitch et al., 2004; Bloom et al., 2010). It is also assumed that photorespiration plays a role in photoprotection under high light and drought conditions (Lima Neto et al., 2017; Bai et al., 2008; Guan et al., 2004). However, the dark side of the photorespiration reaction is that it is indeed a wasteful process. During photorespiration one molecule of NH_3 is also released together with a prefixed CO_2 molecule. Re-assimilation of the NH_3 in the chloroplast also costs a lot of energy. It has been estimated that further re-assimilation of CO_2 and NH_3 molecule cost 12.25 ATP (Peterhansel et al., 2010) in a C_3 plant. Considering the adverse effect of photorespiration, it is assumed that a reduction in the

photorespiration rate would show a considerable promise to enhance the efficiency of photosynthesis, and thereby growth and yield in the C_3 plants. Many research groups are trying to minimize the effect of photorespiration in order to increase the efficiency of photosynthesis, especially in the crop plants. One strategy would be to short-circuit the photorespiratory pathway in plants to achieve recycling of 2-PG without the release of NH_3. A study performed by Kebeish et al., (2007) reported that an introduction of the bacterial glycerate pathway in the plant chloroplasts would bypass the photorespiration reaction to some extent. The authors targeted the five key catabolic enzymes of *Escherichia coli* glycerate pathway into the chloroplasts of *Arabidopsis thaliana*, with the aim to convert the photorespiratory glyoxylate into glycerate. It was proposed that the introduction of the bacterial glycerate pathway in plant chloroplasts would reduce the formation of photorespiratory glycine and its conversion of serine, which is accompanied by a release of NH_3. The transgenic plants produced in that experiment were indeed showed a reduced rate of photorespiration and increased biomass yields under short day condition. Later, similar types of experiments were also conducted in other C_3 plants (Dalal et al., 2015; de FC Carvalho et al., 2011); in most of the cases higher photosynthesis rates and yields were obtained. However, none of the studies showed qualitative and quantitative evidence for the claim that the bacterial glycerate pathway truly functions in the *planta* and the anticipated bypass of the photorespiration has occurred.

Another strategy to reduce the photorespiration would be to improve the catalytic activity of Rubisco with better affinity for CO_2 over O_2 (Marcus et al., 2011; Parry et al., 2012). To the best of our knowledge there has been only little success so far for engineering an improved Rubisco due to its complex structure (Carmo-Silva et al., 2015). However, nature has already evolved a solution over the oxygenation problem of Rubisco. To suppress the photorespiration reaction, in addition to the C_3 cycle, some plants have evolved a new elaborate cycle of photosynthesis known as C_4 photosynthesis.

2. C_4 photosynthesis: An evolutionary adaptation against the photorespiration problem

2.1. CO_2 concentration mechanism of C_4 plants relies on specialized leaf anatomy

The C_4 photosynthesis pathway serves as a "CO_2 pump" to concentrate atmospheric CO_2 surrounding the Rubisco, leading to the reduction in oxygenase reaction and, accordingly, photorespiration. Though a few C_4 species have been reported to perform C_4 photosynthesis within a single cell (Edwards et al., 2004), however, the photosynthetic pathway in most of the C_4 plants is compartmented into two different cell types viz., the mesophyll cell (M) and the bundle sheath cell (BS). These two cell types are arranged in a wreath-like structure around the vascular bundle, which is called the Kranz anatomy, first described by an Austrian botanist named Haberlandt (1881). The M cells are located toward the outer pace of the leaf and so are in contact with intercellular airspace, while the BS cells are arranged internally in the M cells and so are close to the vascular bundle (Dengler & Nelson, 1999). In the majority of C_4 plants, M cells are typically arranged in a radial direction, an arrangement that allows each cell to be in contact with the BS cells. The volume of the intercellular spaces in C_4 plants is also reduced when compared with the C_3 plants. In addition to that, interveinal distance and the leaf thickness of C_4 plants is often reduced in counterparts to their C_3 relatives. To carry out the complex process of photosynthesis, the C_4 plants often adapted larger BS cells with more chloroplasts and other cell organelles in them (Ueno et al., 2006; Muhaidat et al., 2011; Lundgren et al., 2014). In C_4 plants, an almost equal number of chloroplasts could be found between the M and BS cells. In contrast, the C_3 plants harbor 90 % of their total chloroplasts in the M cells, as the entire photosynthesis of C_3 plants takes place in the M cells (Kinsman & Pyke, 1998; Leegood, 2008) (Figure 3). Furthermore, in many of the C_4 plants (i.e., NADP-ME types of C_4 plants), the outer layer of the BS cell wall is impregnated with suberin in order to prevent the escape of CO_2 (Brown, 1975; Rao & Dixon, 2016). Among the C_4 species without a suberin layer in the BS cell wall, there is a tendency to locate their chloroplasts on the inner side (centripetal) of the wall, and the large vacuole of the BS cell helps to slow down the leakage of CO_2 (Sage, 2004). To fulfill the requirement of C_4 cycle, photosynthetic metabolites must be exchanged between the two cell types. Numerous plasmodesmata connecting the M and BS cells in C_4 plants are thought to provide a pathway for symplastic diffusion of photosynthetic metabolites between the two cell types (Botha,

1992). In comparison with a C_3 plant, a C_4 plant often contains a higher density of plasmodesmata between the M and the BS cells junction (Botha, 1992; Danila et al., 2016). A recent study showed that the higher plasmodesmata densities in the C_4 plants are achieved by either increasing the number of pit-field areas or by the number of plasmodesmata per pit-field area (Danila et al., 2018; Danila et al., 2016). A cryptic Kranz-anatomy structure can also be observed in many of the C_3 species (Kinsman & Pyke, 1998); however, their function and role had to be changed to fulfill the role of C_4 photosynthesis.

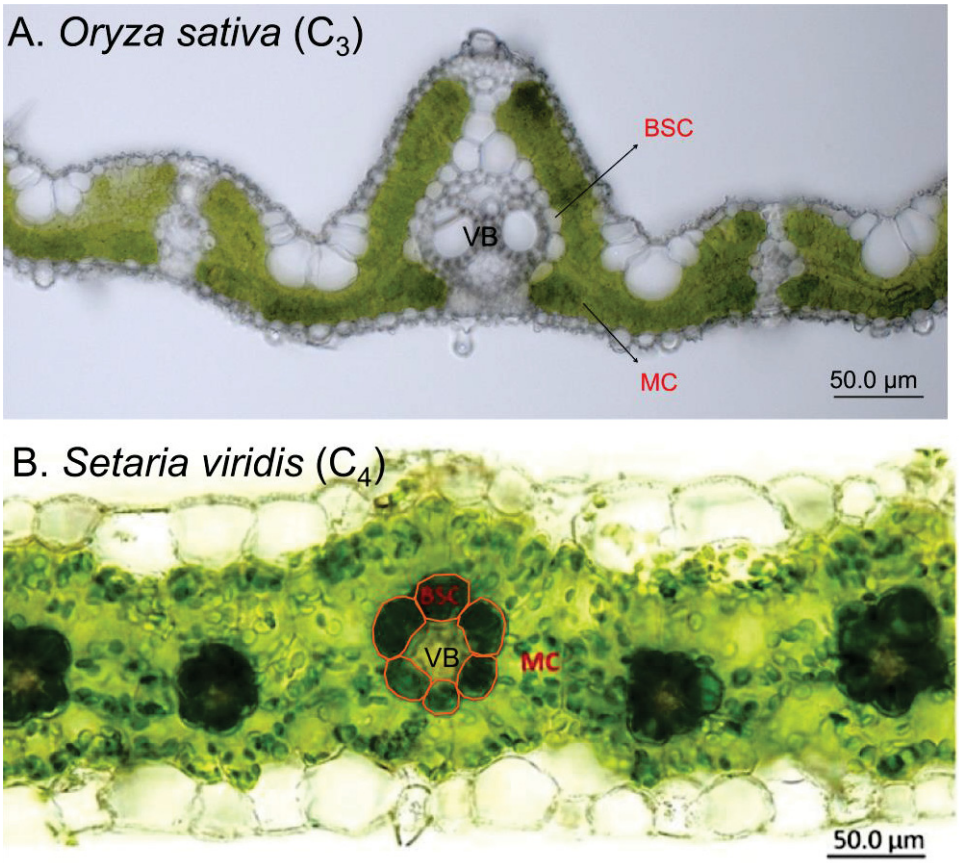


Figure 3: A comparison between a C_3 leaf (*Oryza sativa*) and a C_4 leaf (*Setaria viridis*) anatomy. C_4 plants often contain a higher M and BS cells ratio in comparison to the C_3 plants. The BS cells of the C_4 plants are larger than the C_3 plants and have more chloroplasts and other cell organelles compared to C_3 plants. Here, MC, Mesophyll cell; BSC, Bundle sheath cell; and VB, Vascular bundle. The BS cells of *Setaria viridis* are marked with the orange circles. The leaf cross-section of *Setaria viridis* was taken from (Karki et al., 2013).

2.2. C_4 photosynthesis cycle: A metabolic cooperation between M and BS cells

The two-celled C_4 photosynthesis cycle starts in the cytosol of the M cells. Atmospheric CO_2 is initially hydrated to bicarbonate (HCO_3^-) by carbonic anhydrase. The bicarbonate is then fixed by phospho*enol*pyruvate carboxylase (PEPC, EC 4.1.1.31) into the substrate phospho*enol*pyruvate (PEP), which leads to the formation of a four-carbon compound, oxaloacetate (OAA). PEPC has no substrate affinity for O_2; therefore, it can perform carboxylation reaction under an environment of high O_2 concentration. Besides, PEPC can fix the atmospheric CO_2 much faster than Rubisco (von Caemmerer, 2010). The four-carbon acid OAA so produced is the first stable compound of this pathway, and, so, the photosynthesis cycle is referred to as C_4 photosynthesis. Depending on the type of C_4 photosynthesis, the OAA can be reduced into malate or transaminated to aspartate. The resultant C_4 dicarboxylic acids are then diffused into the BS cells chloroplasts through plasmodesmata. In the BS cell, the C_4 acid is then decarboxylated by the decarboxylation enzyme, and released CO_2 surrounding the Rubisco, and subsequently metabolized by the Calvin–Benson cycle. The decarboxylation of C_4 acid can take place through any of the three decarboxylation enzymes: NADP-malic enzyme (NADP-ME, EC 1.1.1.40), NAD-malic enzyme (NAD-ME, EC 1.1.1.39), or PEP-carboxykinase (PEPCK, EC 4.1.1.49). These decarboxylation enzymes give the major subtypes of C_4 photosynthesis. Although the C_4 plants are assigned to one of these three decarboxylation reactions, recent molecular, biochemical, and physiological data suggest that flexibility is present between the decarboxylation pathways. Many C_4 plants have been found to use more than one decarboxylation enzyme (Furbank, 2011; Bräutigam et al., 2014; Wang et al., 2014). For example, *Zea mays,* which is a typical NADP-ME type plant, has also been found to perform additional decarboxylation of aspartate by PEPCK (Furbank, 2011). In the final step of C_4 pathway, the remaining three-carbon compound (Pyruvate or Alanine) of the decarboxylation reaction returns to the M cells, where the PEP is regenerated to maintain the C_4 cycle in a continuously active manner.

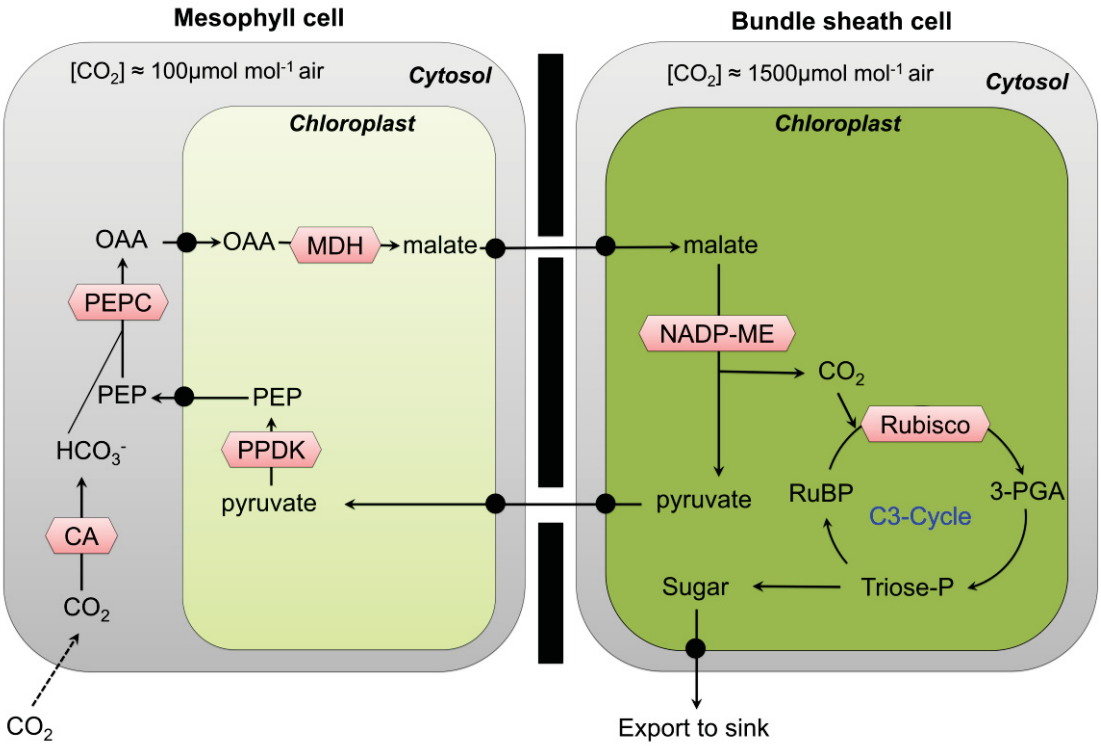


Figure 4: An NADP-ME type of C_4 photosynthesis pathway. The complex process C_4 photosynthetic pathway is achieved through metabolic cooperation of mesophyll and the bundle sheath cells. Enzymes involved in this pathway are represented in hexagonal structure. Here, CA, Carboxy anhydrase; PEPC, Phospho*enol*pyruvate carboxylase; MDH, Malate dehydrogenase; NADP-ME, NADP-malic enzyme decarboxylase; Rubisco, Ribulose bisphosphate carboxylase oxygenase; PPDK, Phospho*enol*pyruvate di-kinase. The metabolites are OAA, Oxaloacetate, PEP, Phospho*enol*pyruvate); 3-PGA, 3- Phosphoglyceric acid.

In contrast to C_3 plants, Rubisco in C_4 plants is only located in the BS cells (Hatch, 1987). The pre-fixation of inorganic CO_2 in the M cells by the PEPC and decarboxylation of the C_4 acids in the BS cell allows a biochemical "CO_2 pump". This leads to an elevated CO_2 level within the BS cells and allows complete saturation of the active site of Rubisco with the CO_2. Therefore, the oxygenase reaction by the Rubisco in the C_4 plants is drastically reduced, as also the photorespiration (Sage, 2004; Furbank, 2011). In comparison with the C_3 plants, the photorespiration reaction in the C_4 plants is almost undetectable. However, photorespiration reactions in the C_4 plants are not completely abolished. It has been reported that a knock-out of the photorespiratory glycolate oxidase (*GOX*) gene in the C_4 plant maize causes a lethal phenotype in an ambient air condition (Zelitch et al., 2009). The lethal phenotype of

the mutant could be rescued in a high CO_2 condition, suggesting the potential importance of photorespiration in C_4 plants (Somerville, 2001). Since no other metabolic pathway exists in the plant to remove the glycolate, the C_4 plants must carry on the photorespiration reaction in order to prevent the accumulation of this toxic metabolite.

Enhanced photosynthesis efficiency in the C_4 plants not only increases the productivity (biomass yield) but is also associated with the water and nitrogen (N_2) use efficiency. The biochemical CO_2 pump allows the C_4 plants to perform a high rate of photosynthesis even at a low CO_2 concentration in the intercellular airspace of the leaf. This helps the C_4 plants limit the opening of their stomata when compared with the C_3 plants, and thereby prevent loss of water through transpiration (Sage, 2004). In addition, the C_4 plants require three to six times less Rubisco than C_3 plants (Ku et al., 1979; Sage et al., 1987). Since the Rubisco constitutes 10–25 % of the total leaf N_2 in the C_3 plants (Furbank, 2011; Carmo-Silva et al., 2015), a less amount of Rubisco in the C_4 plant implies a higher N_2 use efficiency (Ehleringer & Monsoon, 1993; Oaks, 1994; Long, 1999). Moreover, the Rubisco from C_4 plants is much more improved; it has a better catalytic turnover rate compared with their C_3 relatives (Carmo-Silva et al., 2015). All in all, the C_4 photosynthetic plants show better photosynthetic, water, and nitrogen use efficiency than the C_3 species under conditions of high lights, hot temperatures, and drought.

2.3. Evolution of the C_4 photosynthesis: Transition of C_3 to C_4 photosynthesis

Despite its complexity, the C_4 photosynthesis has evolved no less than 66 times from their ancestral C_3 plants (Sage et al., 2012). The C_4 flora is only distributed in the angiosperm within more than 418 genera, including 8,145 species (Sage, 2016). All these species are distributed in at least 19 families of mono and dicotyledon plants. However, these families are phylogenetically distant from each other, indicating that the C_4 plants must have evolved independently from their C_3 ancestors. Most of the C_4 plants occur in the grasses with 20–23 distinct lineages comprising 5,044 species (Sage, 2016). Also, all the grass C_4 lineages occur within the branch of the family known as PACMAD clade. The PACMAD clade is named after the first initial of the subfamilies'

names, including: Panicoideae, Arundinoideae, Chloridoideae, Mircrairoideae, Aristidoideae, and Danthoideae. Apart from the grass lineages, six lineages in the sedges (Cyperaceae) comprise 1,322 C_4 species (Besnard et al., 2009; Roalson et al., 2010; Sage, 2016), nine lineages in the Chenopodiaceae contain 558 C_4 species (Kadereit et al., 2012; Sage, 2016), and approximately 1,750 C_4 species can be found in the eudicots with more than 34 *de novo* origins (Sage, 2016). Despite being used only by 3 % of the world's total plant species, the C_4 photosynthetic pathway accounts for ~25 % of the total terrestrial biomass production.

It is widely accepted that the C_4 photosynthesis is relatively new in evolutionary terms (Ehleringer et al., 1997). All the C_4 lineages are assumed to have evolved around or after the Oligocene drop in atmospheric CO_2 (Sage et al., 2011; Aliscioni et al., 2012). The most primitive C_4 origin can be found in the grass subfamily Chloridoideae, which evolved during the mid-Oligocene (approximately 30 Mya). In terms of evolutionary events, the Asterecean *Flaveria* is considered the youngest C_4 lineage, and is estimated to have evolved during the last five million years. This C_4 lineage contains a set of species of true C_3, C_3-C_4 intermediates, C_4-like, and true C_4 plants (Edwards & Ku, 1987; McKown et al., 2005). It has been found that there are nine C_3-C_4 intermediate species in the *Flaveria* genus. Owing to its large collection of C_3-C_4 intermediates, *Flaveria* genus has become a role model for studying C_4 evolution.

In earlier, it was proposed that low CO_2 concentration in the environment triggers C_4 evolution. However, the recent studies suggest that the low CO_2 concentration is only a precondition for C_4 evolution. Several environmental factors like heat, salinity, high light, and ecological disturbances are thought to encourage the evolution of C_4 photosynthesis. These environmental factors can also lower the CO_2 concentration in the C_3 plants, thereby favoring the condition for photorespiration. The elevated level of photorespiration in leaves is assumed to be the driving force for the emergence of C_4 photosynthesis. Careful investigations of evolutionary patterns of many C_4 species have led to a pyramid-like model for C_3 to C_4 transition (Sage, 2004; Sage et al., 2012). According to the proposed model, the gradual evolution of C_4 traits occurs from general preconditioning to integration and optimization of the final trait.

The road from C_3 to C_4 transition starts with the preconditioning steps. This includes

duplication of the relevant genes so that multiple copies of a single gene can exist in plants. This helps the plant experiment with the copies of the gene for neo-functionalization without losing the function of the original gene (Lynch & Conery, 2000). In the next step of C_4 evolution, anatomical preconditioning of leaf occurred. At this stage, the distance between the M cells and BS cells get reduced, allowing efficient diffusion of C_4 metabolites between the two cell types (Ehleringer, 1997). This is achieved by reducing the interveinal distances and enlarging the BS cells. Denser vein spacing has been shown to supply sufficient water to avoid desiccation in the M cell with open stomata in hot and arid environments (Brodribb & Feild, 2010). Moreover, it is assumed that a reduction in the interveinal density of the leaf also reduces the evaporative surface area and the enlarged BS cells enhance water storage in hot and dry conditions. Hence, a closer vein and enlarged BS cells are an adaptation to warm and dry conditions (Scoffoni et al., 2011; Sage et al., 2012). Moreover, the increased vein density also provides a mechanical support to the plant in windy and hot climate locations.

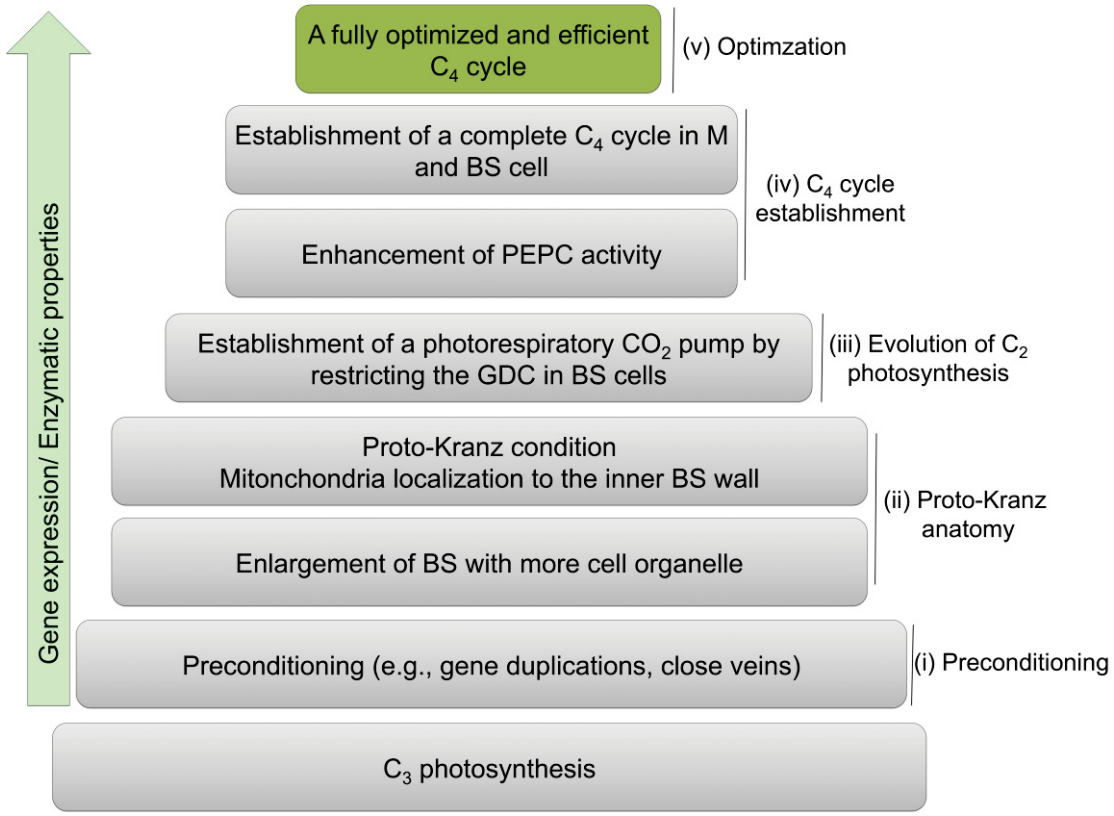


Figure 5: The conceptual pyramid-model for the progressive evolution of C_3 to C_4

photosynthesis. The model represents the five major phases of the C_3 to C_4 photosynthesis transition. Here, BS, Bundle sheath cell; M, Mesophyll cell; GDC, Glycine decarboxylase complex; PEPC, Phospho*enol*npyruvate carboxylase. Figure adapted from the (Gowik & Westhoff, 2011; Sage et al., 2012).

Once the criteria for the anatomical preconditions are fulfilled, plants enter the next step of the C_4 evolution. This is the establishment of a so-called "Proto-Kranz" anatomy. This step includes photosynthetic activation of the BS cells by increasing the number of chloroplasts and mitochondria in it. Besides, it can be observed that the mitochondria in the BS cells are arranged in a centripetal orientation, resulting in the establishment of a one-cell glycine shuttle system (Muhaidat et al., 2011). This allows plants to develop a condition for a photorespiratory CO_2 pump. At this stage, plants are ready to enter the most critical step of C_4 evolution viz. the development of C_2 photosynthesis.

Once the protokranz anatomy has developed, it influences the relocation of photorespiratory glycine decarboxylase complex (GDC) from the M cell into BS cells. This relocation of the GDC causes the splitting of photorespiration reaction between the M and BS cells. During the photorespiration reaction, GDC complex converts the photorespiratory glycine into a serine by releasing one molecule of CO_2. A loss of GDC in the M cell also reduces the photorespiratory CO_2 loss in this tissue because glycine can no longer be decarboxylated in the M cells. In order to prevent the accumulation of photorespiratory products in the M cells, photorespiratory glycine must be transported into the BS cells. Hence, a constant flow of glycine from the M cells to the BS cells take place, and is decarboxylated by the BS cell-specific GDC. In this way, CO_2 lost from the photorespiration reaction became a new source of CO_2 for the chloroplast in the BS cells (Bauwe, 2010). Shifting of the GDC into the BS cells thereby creates a photorespiratory CO_2 pump in the BS cells (Figure 6). This phenomenon can be observed in many of the C_3-C_4 intermediate species, including the genera of *Flaveria* (Monson et al., 1984; Hylton et al., 1988) and *Heliotropium* (Muhaidat et al., 2011).

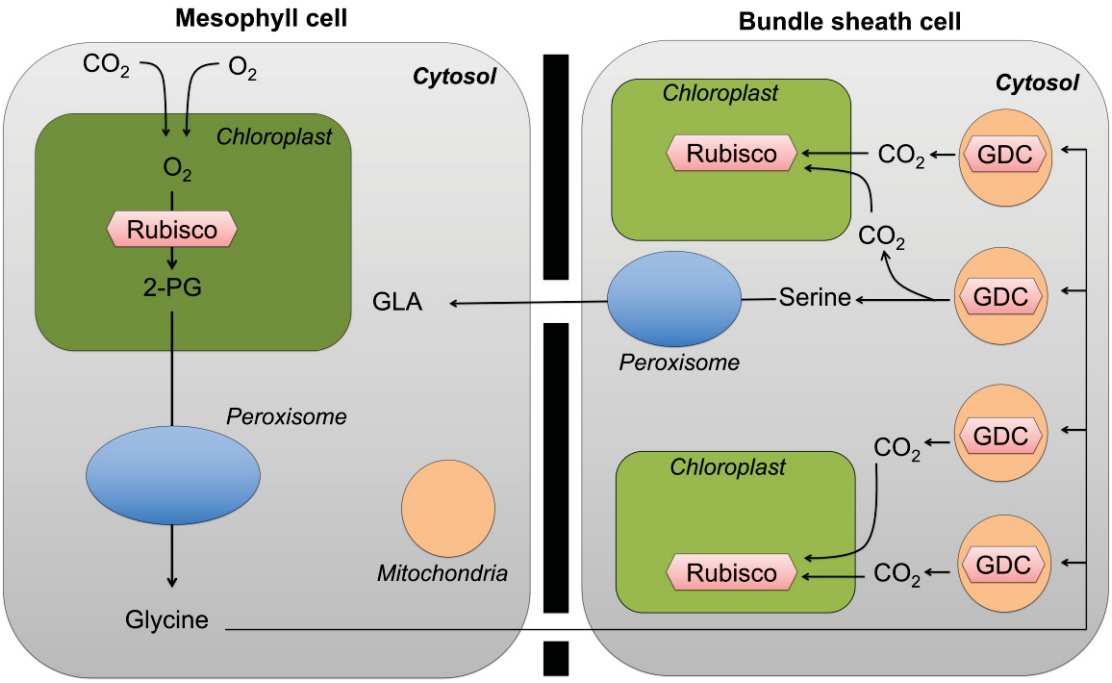


Figure 6: A simplified photorespiratory C2 photosynthesis cycle. Translocation of the glycine decarboxylase complex (GDC) into BS cells mitochondria creates a constant flux of photorespiratory glycine from M cell to the BS cells. Decarboxylation of the photorespiratory glycine in the BS cells leads to CO_2 enrichment in this tissue and suppresses the oxygenase activity of Rubisco. Enzymes involved in this pathway are represented in hexagonal structure. Here, PG, Phosphoglycolate; GLA, Glycerate. This figure is adapted from (Sage et al., 2012).

The fourth phase of the C_4 evolution is dedicated to the establishment of a complete C_4 cycle in the M and BS cells. This includes enhanced and strict compartmentalized expression of the key C_4 enzymes in between the M and BS cells. For example, the two main carboxylating enzymes PEPC and Rubisco are expressed in the M and BS cells respectively. It has been verified in many studies that differential expressions of the C_4 genes are mainly controlled at multiple levels, including transcriptional, post-transcriptional, post-translation, and epigenetic (Li et al., 2010; Wiludda et al., 2011). It has been reported that even a single nucleotide change in the *cis* regulatory element can lead to M cell-specific expression of a Phospho*enol*pyruvate carboxylase gene (*Ppc*) (Gowik et al., 2004). This indicates the evolution of a C_4 gene from its C_3 ancestor may not require drastic changes in genetic terms.

Finally, a fully functional C_4 cycle is established with a fine-tuning of the C_4

enzymes kinetics. It has been found that the C_4 cycle enzymes differ from the C_3 isoforms in terms of kinetic and regulatory characteristics (Ting & Osmond, 1973; Svensson et al., 2003). For this, coding sequences of C_4 enzymes had to be changed in an appropriate manner to adapt the C_4 characteristic kinetic properties.

3. Phospho*enol*pyruvate carboxylase: The entry enzyme of the C_4 photosynthesis pathway

3.1. Functions of the phosphoenolpyruvate carboxylase in plants

Phospho*enol*pyruvate is a cytosolic enzyme and catalyzes the irreversible carboxylation of phospho*enol*pyruvate (PEP) in the presence of Mg^{2+} and bicarbonate (HCO_3^-) to yield a C_4 acid, oxaloacetate. In plants, four different isozymes of PEPC have been recognized: C_4 type, C_3 type, root type, and bacterial type (Ku et al. 1996; Masumoto et al., 2010). Only the C_4 type PEPC serves initial fixation of CO_2 in the C_4 pathway and is exclusively expressed in the M cells cytosol of C_4 leaves. The photosynthetic C_4 isoform of *Ppc* gene has evolved independently several times from the ancestral C_3 non-photosynthetic isoforms (Kellogg, 1999). Non-photosynthetic isoforms of *Ppc* with different catalytic and metabolic activities are found in both photosynthetic and non-photosynthetic tissue of all plants (Latzko & Kelly, 1983). In comparison with the high expression of *C_4Ppc* in the M cells, the non-photosynthetic *Ppc* isoforms are expressed moderately in all plant tissues (Hermans & Westhoff, 1990; Ernst & Westhoff, 1997). In bacteria, algae, and higher plants, the non-photosynthetic PEPC isoforms are involved in a variety of metabolic processes, including anapleotropic CO_2 fixation, production of carbon skeletons in nitrogen fixation, maintenance of ion balance, modulation of turgor in stomatal guard cells, pH regulation, and others (Ting & Osmond, 1973a; Winter, 1985; Melzer & O'Leary 1987; Schuller et al., 1990; Cushman & Bohnert, 1999). However, C_4 PEPCs have some distinct properties that differentiate them clearly from the other types of PEPCs of higher plants. C_4 PEPCs show a lower affinity to PEP as a substrate compared to the other non-photosynthetic PEPCs (Gowik et al., 2006; Lara et al., 2006). Similarly, C_4 PEPCs bind bicarbonate with a higher affinity than other PEPCs that bind bicarbonate. The differences in the substrate affinity indicate that the C_4 PEPCs acquire the C_4 determinants during the evolution (Gowik & Westhoff, 2011). Furthermore, C_4 PEPCs

are expressed only in the M cells. In contrast to C_4 PEPCs, the C_3 PEPCs are expressed at a much lower level and not in a cell-specific manner (Sheen, 1999). Although a major difference is found in the cell-specific expression of C_3 and C_4 types *Ppc*, the structure of coding sequence is highly conserved between the C_3 and C_4 type *Ppc* (Gowik & Westhoff, 2011). Strong M cell-specific expression patterns of *C_4Ppc* must have occurred during the evolution from the ancestral C_3 isoforms. There is compelling evidence that cell-specific expressions of *C_4Ppc* have evolved by changes in the transcriptional control of C_3 isoforms (Stockhaus et al., 1997; Sheen, 1999; Gowik et al., 2004).

3.2. Mesophyll cell-specific regulatory mechanism of the C_4Ppc gene (ppcA) in genus Flaveria

The *ppc*A gene of C_4 *Flaveria* species was found to encode the C_4 type PEPCs. A similar type of *ppc*A genes can also be found in the C_3 *Flaveria* species, where they encode non-photosynthetic isoforms of PEPC (Bläsing et al., 2002). The *ppc*A coding sequence of C_3 and C_4 *Flaveria* species share a 96 % identical sequence but express differentially (Bläsing et al., 2002). It is assumed that the *ppc*A gene has evolved from the gene duplication event of a *ppc*B gene, and both *ppc*A and *ppc*B gene must have existed in the last common ancestors of the C_3 and C_4 *Flaveria* species (Bläsing et al., 2002). The role of the *ppc*B gene in the plants is not precisely known. However, the kinetic properties and accumulation pattern of the *ppc*B transcripts suggest that it might have a housekeeping gene function (Ernst & Westhoff, 1996; Bläsing et al., 2002). In comparison with the C_4 *Flaveria* species, the *ppc*A of the C_3 *Flaveria* species are found to be expressed very weakly in leaves, stems, and root tissues. However, the exact function of the *ppc*A gene in the C_3 *Flaveria* species is still elusive. One could raise a question then as to what makes the *ppc*A gene mesophyll cell-specific with high expression level in the C_4 *Flaveria* species. Promoter-reporter gene study with the *ppc*A 5′-flanking sequences of the *Flaveria triveria* (C_4) in the transgenic *Flaveria bidentis* (C_4) revealed that the M cell-specific expression is indeed controlled at the transcriptional level (Stockhaus et al., 1997). Further analysis showed that two parts—designated as the proximal region (PR) and the distal region (DR)—mainly contribute to the spatial expression, and the expression strength of the *ppc*A promoter. The PR ranges from -1 to -570 bp; it is responsible for the quantitative expression of the *ppc*A

gene. On the other hand, the DR ranges from -1566 to -2141 and consists of a 41 bp enhancer like expression module known as mesophyll expression module-1 (MEM-1). The MEM-1 module is thought to be responsible for the M cell-specific expression pattern of *ppc*A gene by suppressing its expression in the BS cells. Just the combination of the PR and DR is able to drive strong M cell-specific expression of a reporter gene. A similar homolog of MEM-1 could also be found in the *ppc*A promoter of the C_3 and C_3-C_4 intermediate *Flaveria* species. However, the C_3 and C_4 specific MEM-1 sequences differ at two positions viz. a G to A transition and an insertion of the tetra-nucleotide CACT in the C_4 MEM-1 in comparison with the C_3 one. The two changes in the MEM-1 sequence appear to be mandatory for the evolution of M cell-specific expression of the *C_4 ppc*A gene (Akyildiz et al., 2007; Gowik et al., 2004). In a yeast hybrid assay, it was shown that the zinc finger homeodomain transcription factors FtHB1, 3 and 4 from *Flaveria trinervia* can interact with the *ppc*A 5´-flanking sequences and was proposed to control the M cell-specific regulation of *ppc*A gene in C_4 *Flaveria* (Windhovel et al., 2001). However, the deletion of the potential FtHB proteins binding sites in the *F. trinervia ppc*A 5´-flanking sequences did not show any significant effect on M cell-specific expression pattern (Engelmann et al., 2008). Hence, the *trans* regulatory factors responsible for directing the M cell-specific expression of C_4 *ppc*A gene remain hidden from detection.

3.3. Mesophyll cell-specific regulatory mechanism studied in the grass C_4Ppc gene

Extensive studies were earlier carried out to resolve the M cell-specific regulatory mechanism of the maize *C_4Ppc* gene. The 5´-flanking sequences of the maize *C_4Ppc* gene was found to drive strong M cell-specific expression of a reporter gene in C_3 rice (Matsuoka et al., 1994). This indicates that like the *ppc*A gene of C_4 *Flaveria*, M cell-specific regulatory mechanism of the maize *C_4Ppc* gene is also regulated at the transcriptional level. A comparison of the C_3 and C_4 type 5´-flanking sequences of *Z. mays Ppc*s' revealed that the sequences are identical at a range from the TATA box element to the translational start codon ATG (Schäffner & Sheen, 1992). Diverse sequences upstream of their TATA box suggested that the *C_4Ppc* gene of *Z. mays* might have evolved from an ancestral *C_3Ppc* gene after an unequal recombination event near the TATA box element (Schäffner & Sheen, 1992; Sheen, 1999). This recombination event might have brought all the *cis* elements necessary for the M cell-specific

expression of *C_4Ppc*. An earlier study, performed by Langdale et al. (1991), proposed that M cell-specific expression is mediated by differential methylation of a *Pvu*II site in the far upstream (-3.5 kb) region of *ZmPpc* promoter. Further experiments by Tolley et al. (2012) also provided enough evidence that methylation is not required to direct the cell-specific expression of *ZmPpc*. Yanagisawa & Sheen (1998) proposed that the nuclear zinc finger proteins Dof-1 and Dof-2 play a crucial role for the M cell-specific expression of *ZmPpc*. According to their hypothesis, Dof-1 and Dof-2 proteins are expressed differentially in the M cells and BS cells respectively. The Dof-1 protein is assumed to act as an activator of *C_4Ppc* promoter whereas the Dof-2 protein acts as a repressor. Therefore, there appears to be an action of antagonism between the Dof-1 and Dof-2 proteins. Although later Taniguchi et al. (2000) performed a gel shift experiment that indicated that there was no interaction between the aforementioned Dof transcription factors and the *ZmPpc* 5′-flanking sequences.

In another study, it was reported that the nuclear factor PEP-I could form a DNA-protein complex with the *ZmPpc* 5′-flanking sequences (Kano-Murakami et al., 1991). This DNA-protein complex was only detected with the nuclear crude extract from green leaves but not from roots or etiolated leaves (Kano-Murakami et al., 1991). The nuclear factor PEP-I has been found to interact with the GC rich elements on the *ZmPpc* 5′-flanking sequences. The proposed consensus sequence of the GC rich elements in the *ZmPpc* 5′-flanking sequences is CCCTCTCCACATCC and the CTCC is thought to be essential for binding of PEP-I. It was proposed by Kano-Murakami et al. (1991) that the PEP-I binding site could play a vital role for M cell-specific expression mechanism. However, gel shift analysis performed by Taniguchi et al. (2000) revealed that there are no cell type-specific nuclear factors binding to the PEP-I site. This finding challenges the proposed hypothesis of the M cell-specific regulation by PEP-I site. The authors also reported that the proximal 600 bp of *ZmPpc* promoter is capable of driving M cell-specific expression of reporter gene GUS in the transgenic maize leaf. Further studies revealed that the proximal 600 bp of the *ZmPpc* 5′-flanking could interact with some unidentified nuclear proteins in a cell-specific manner (Taniguchi et al., 2000). According to their findings, the unidentified nuclear factors—termed PEP_{Ib} and PEP_{Ic}—are specific to M cells, whereas the nuclear factors PEP_{Ia} and PEP_{IIa} are specific to BS cells. The PEP_{Ia} and PEP_{IIa} formation in the M cells is extremely inhibited by light. Conversely, the PEP_{Ib} and PEP_{Ic} formation in the M cells is attenuated by light

during the greening of leaves. However, the binding sites of these nuclear proteins remain unknown. Therefore, the molecular mechanisms and the *cis*-regulatory element(s) governing the M cell-specific expression of the *ZmPpc* gene remains poorly understood.

4. Improvement of photosynthesis in rice: The C_4-rice project

Global population is increasing day by day and is expected to reach nearly 10 billion by 2050 (United Nations, 2015). To provide adequate food and nutrition to the burgeoning population, agricultural food production needs to be boosted. Rice is a C_3 crop plant majorly grown in Asia. People in Asia mostly depend on rice for their calorific intake. More than 60 % of the world's population is currently living in Asia, where most of the world's existing poverty is also concentrated. Each hectare of land used for rice production currently provides food for 27 persons. However, by 2050, each hectare of land will have to produce rice in an amount that can support food for 43 persons (Sheehy et al., 2008). Therefore, in the near future, Asia must increase rice production by up to 60 % to provide adequate nutrition to its growing population (Sheehy et al., 2008). However, decreasing farmland areas as a result of industrialization and unpredictable climate changes are constantly challenging the agricultural food production. Moreover, the current breeding process would likely be insufficient to increase the rice production by 60 % in 2050. Hence, a "second green revolution" is required in world agriculture to boost rice production with less fertilizer and water uses. This can be achieved only if the photosynthesis efficiency in rice is substantially improved by using solar energy and atmospheric CO_2 more efficiently. As the C_4 photosynthetic plants simply outperform the C_3 plants in hot and arid climates due to high radiation use efficiency (RUE), installing a C_4 engine in C_3 rice is assumed to increase the photosynthesis rate by 50 % and thereby the biomass production (Mitchell and Sheehy, 2006; Zhu et al., 2010). Some attempts to install a single cell C_4 photosynthesis system in rice have already been made (Taniguchi et al., 2008; von Caemmerer et al., 2014). The single cell C_4 photosynthesis can be found in the aquatic plant *Hydrilla verticillata* (L.f) Royle, in which intracellular compartmentalization of enzymatic activities enable it to capture CO_2 at the one end of a cell and decarboxylation and refixation of CO_2 by Rubisco at the other extreme (Edwards et al., 2004). It was

initially hypothesized that if the four major C_4 specific enzymes (PEPC, PPDK, NADP-MDH, and NADP-ME) can be overproduced in rice leaves, then the C_4-like pathway of *Hydrilla verticillata* might also operate in rice. The four major enzymes that are involved in this pathway were overexpressed in the transgenic rice leaves. However, the transformants did not show any potential improvement in the photosynthesis rate in comparison with their wild type plant (Taniguchi et al., 2008). The failure of this attempt was assumed to be caused by no changes in leaf anatomy, missing appropriate transporters for metabolic pathway, and the fact that the introduced C_4 genes were not appropriately expressed in a cell-specific manner (Miyao et al., 2011). In addition to that, one of the novel problems for engineering a single-cell C_4 pathway into rice is the presence of an endogenous chloroplastic PEPC, which might interfere with the potential C_4 cycle in the M cell of rice (Masumoto et al., 2010; Miyao et al., 2011).

In conclusion, the outcomes of engineering the single-cell C_4 photosynthesis system in rice were not convincing as anticipated and must be reviewed in detail. However, the two-celled C_4 photosynthetic pathway—as established in most C_4 plants—might be a better option for increasing photosynthesis efficiency in rice. For this purpose, the 'International C_4 Rice Consortium' was founded in 2008; it is led by the International Rice Research Institute (IRRI) with 24 participating research groups. The main aim of this 'C_4 Rice Consortium' is to test the feasibility of installing a NADP-ME type C_4 photosynthetic biochemical pathway into C_3 rice (http://irri.org/c4rice). There is no doubt that installing a fully functional C_4 engine into C_3 rice is an enormous challenge and requires many small and large changes. Based on the current knowledge of the C_4 photosynthesis mechanism, the following engineering steps at least must be implemented into rice to develop a fully functional C_4 pathway in rice leaf.

(i) Introduction of higher order of veins to reduce the vein spacing, thereby increasing the vein density in rice leaf.

(ii) The typical Kranz anatomy of a C_4 leaf often consists of a repeating pattern of vascular bundle-BS-M-M-BS. Therefore, in rice leaf, the BS to M cell ratio must be increased.

(iii) In rice, 90 % of the total chloroplasts are in the M cells (Ueno et al., 2006). In contrast, the C_4 plants contain an almost equal number of chloroplasts in between the M and BS cells. Because of that, the C_4 plants adapted enlarged

BS cells with more chloroplasts and other cell organelles. In addition to that, the BS cells are interconnected with the neighboring M cells through plasmodesmata. To install the C_4 engine into rice, the BS cells of rice leaves must be activated and adapted like C_4 plants.

(iv) The BS and M cells chloroplasts of most C_4 plants are dimorphic in nature (Majeran et al., 2010). For example, BS cells' chloroplasts accumulate starch and are distributed centrifugally. In addition to that, BS cells chloroplasts often display abundant thylakoids, which do not associate into grana. In contrast, the M cells chloroplasts of C_4 plants are similar to that of the C_3 plants chloroplasts but do not accumulate starch grains (Majeran et al., 2010). Therefore, the engineering of the C_4 type dimorphic chloroplasts in the M and BS of the rice leaves would be a crucial step for C4 rice development.

(v) Finally, the activity of the Calvin–Benson cycle in the M cells of rice leaves should be decreased and must be enhanced in the BS cells. Besides, the photorespiration reaction rate in the M cells has to be decreased, which can be achieved by the restriction of the Rubisco and the Calvin cycle enzymes in the BS cells. In order to do that, the key C_4 enzymes must be expressed in a cell-specific manner—either in the M cells or in the BS cells (Hibberd & Covshoff, 2010).

All the genes involved in the two-celled C_4 photosynthesis pathway already existed in the ancestral C_3 plants. The C_3 isoforms of these C_4 enzymes do not play a major role in leaf photosynthesis; rather, they carry out the other metabolic functions within the C_3 plants, e.g., carbohydrate and nitrogen metabolism (Aubry et al., 2011; Brown et al., 2011). Moreover, the C_4 plants have evolved independently 66 times in at least 18 different plant families, indicating that the evolution of C_4 plants from their C_3 ancestor plants might be relatively easy in genetic terms (Sage, 2016). This suggests that switching from the default C_3 photosynthetic mechanism into C_4 photosynthesis does not require the introduction of new players, but rather changes in regulation, kinetics, and tissue specificity of existing enzymes and their respective genes (Doebley & Lukens, 1998; Sage, 2004; Schuler et al., 2016). Working groups of the C_4 rice consortium from all over the world are trying to discover the genes related to the development of Kranz anatomy and the C_4 biochemical process. Once tested, the genes

regulating the C_4 leaf anatomy and biochemical processes will be introduced into the rice plants. For this, the expression of several genes has to be redirected to either mesophyll (M) or bundle sheath (BS) cells. Therefore, a set of cell-specific promoter units are highly required, which can be used alternatively for the generation of C_4 rice. To the best of our knowledge, only two promoters (promoters of *C_4Ppc* and *PPDK* genes of maize) have been reported and extensively used so far—these confer M cell-specific specificity in the rice leaf tissue at the level sought (Matsuoka et al., 1994, Matsuoka et al., 1993). In contrast, *Zoysia japonica PCK* promoter is the only reported promoter that produces BS cell-specific expression in rice plant. The number of available cell-specific promoters for rice might be insufficient for the C_4 rice project. Moreover, the utilization of the same promoter for multiple genes transfer in rice might also cause a silencing effect because of repetitive elements. It is necessary, therefore, to explore more cell-specific promoters for the sake of C_4 rice project. Consequently, an aim of this thesis is to investigate a new set of M cell-specific promoters for the rice leaves.

5. An EMS-based forward genetics approach to identify the genes for photosynthetic activation of BS cells in C_3 plants

Photosynthetic activation of the BS cells is one of the most important steps during the transition from C_3 to C_4 evolution. In contrast with the C_3 plants, the C_4 plants exhibit enlarged BS cells with more chloroplasts and other cell organelles. In order to install a fully functional C_4 cycle in the C_3 crop plants, such as rice, the BS cells must be activated in a C_4 manner. However, the genes regulating the activation of BS cells and the development of Kranz anatomy are still hidden from detection.

One of the first reported transcription factors, which plays a vital role in the development of BS chloroplasts, was the maize GOLDEN2 (G2) protein. A loss-of-function mutant of *G2* allele in maize showed a retarded chloroplast development in the BS cells, but the M cell chloroplasts remained unaffected (Hall et al., 1998; Waters et al., 2008; Waters & Langdale, 2009). Further, a transcriptomics study with the maize leaf tissue revealed that *ZmG2* and its paralog *GOLDEN2-LIKE* (*ZmGLK1*) are

expressed predominantly in the BS and M cells respectively (Li et al., 2010). Overexpression of the *ZmGLK1* and *ZmG2* in C_3 rice has been reported to show an increase in chloroplast size in the BS cells (Wang et al., 2017). This implies that *GLK* might be an important regulatory factor for inducing C_3 to C_4 characteristics. Another transcriptional regulatory module, the so-called SHORTROOT-SCARECROW, was also found to regulate the establishment of Kranz-anatomy in the maize leaf tissue (Sparks et al., 2016; Slewinski, 2013). Nevertheless, more information is required about the genes that are associated with the BS cell size and BS cell chloroplast development in the C_4 plants. Once the information of the gene is obtained, it could be tested in rice to introduce a C_4-like anatomy in rice leaf tissue.

When the gene information for a particular genetic mechanism is insufficient, then the forward genetics approach with large-scale screening is possibly the best option to study them. It is already accepted that the BS cell is an ancient evolutionary invention in the angiosperms (Sage et al., 2012). In contrast with the C_4 species BS cells, the corresponding tissues in C_3 species are relatively smaller with fewer chloroplasts and other cell organelles. This implies that the BS cells in C_3 plants do not play a major role in the photosynthesis pathway. The exact physiological role of the BS cells in C_3 plants is not precisely known, but is assumed to perform other functions, like phloem loading and unloading. Or it may provide mechanical strength within the leaf (Kinsman & Pyke, 1998; Griffiths et al., 2013). However, the cross-species experiments between the C_3 model plant *Arabidopsis thaliana* and the Asteracean C_4 plant *Flaveria trinervia* revealed that the gene regulatory networks of the BS cells in the C3 and C_4 dicotyledons plants are at least partly conserved (Westhoff & Gowik, 2010). The C_3 model plant *Arabidopsis thaliana* can thereby be used for forward genetics screening in order to discover the genes regulating BS cell ontogeny and maintenance in C_4 plants. The principle of forward genetics screening is to randomly mutate the genes of a whole plant genome to generate the mutants with a phenotype of interest. Later, the genes related to the phenotype of interest are identified through different methods, depending on the mutagenesis types. Two different methods of mutagenesis are currently available to study forward genetics screening in the plant genome. These two methods are: (i) insertional mutagenesis by transposons or T-DNA, and, (ii) induced mutagenesis, either by chemical mutagens or by radiation.

Ethyl methanesulfonate $C_3H_8SO_3$ (EMS) is a chemical mutagen commonly used in the induced mutagenesis application. EMS is an effective and reliable mutagen as it possesses the capability to induce a high mutation frequency in the plant genome. Furthermore, it is relatively easy to use in Arabidopsis. EMS creates mutation in the genetic material by a substitution process, especially by alkalyting the guanine base. The ethyl group of the EMS reacts with a guanine base of DNA and produces an abnormal base, the O^6-ethylguanine. During the DNA replication, the O^6-ethylguanine makes a pair with a thymine base instead of cytosine. Following the next round of the replication process, the G:C pair turn into an A:T (Greene et al., 2003; Till et al., 2007). This transition of a G:C into an A:T leads to a single nucleotide polymorphism (SNP) in the genome (Kim et al., 2006). EMS mutagenesis can produce a few thousand SNPs per individual. These SNPs that are produced can ultimately generate synonymous or non-synonymous mutations in a coding region of a gene. The synonymous mutation changes one codon of an amino acid into another codon for that same amino acid. Often, the synonymous mutation in the coding sequence of a gene has no effect at all. However, the synonymous mutation can produce a substantial impact in the promoter region of a gene by altering the regulatory elements positively or negatively. It may also modify the RNA splicing site, thereby creating an aberrant splicing or altered mRNA stability (Sikora et al., 2011). For that reason, synonymous mutation is termed a silent mutation. On the other hand, the non-synonymous mutations affect the coding sequence of a gene by changing the amino acid at the site of the mutation, or even by inducing a premature stop codon. There are several advantages of the EMS mutagenesis technique over the other mutagenesis approaches. First, EMS mutagenesis can produce a high density of non-bias mutations in the plant genome, which permits saturation of mutagenesis without screening a large number of individual mutant lines. Secondly, in addition to a loss-of-function mutation, EMS can also result in an allelic series, e.g., a strong, intermediate and weak allele for a given gene (Bowman et al., 1991).

Although the EMS mutagenesis provides many advantages for the collection of desired mutant lines, isolation of the corresponding genes related to the phenotype of interest is still a challenging task. Unlike the insertional mutants, EMS mutant lines do not have any molecular tag that would allow the direct cloning of the gene of interest. Isolation of a gene corresponds to a phenotype of interest is mostly done by map-based cloning, often known as positional cloning. Two different resources are required for the

positional cloning. The first one is a molecular or genetic marker, the chromosomal map positions of which are already known. The second resource is the mapping population that is often generated by a crossing of the mutant plants in a specific ecotype (e.g., Columbia) to wild-type plants of another ecotype (e.g., *Landsberg erecta*). The outcrossing of the mutant lines allows the identification of the markers that are closely linked to the mutation of interest. In recent studies, it has been shown that backcrossing of mutant lines can also produce enough genetic diversity to identify the causative SNP (Abe et al., 2012; James et al., 2013). The complete path from a screening of a mutant of interest to the identification of the related SNP appears to be a very lengthy process. However, new advantages in map-based cloning procedure, such as a combination of bulk segregate analysis and whole-genome re-sequencing (SHOREmap), have been shown up to speed up this process to a large extent. Hence, the EMS mutagenesis could be a powerful tool for discovering the genes regulating the development of C_4-like BS cells in Arabidopsis.

II. Scientific aims

In environments where photorespiration is high, C_4 photosynthetic plants have higher efficiency in the use of water, nitrogen, and radiation than the C_3 plants (Ehleringer et al., 1997). It is an ambitious goal to install this superior type of photosynthesis pathway in C_3 crop rice and could possibly boost up its potential yield. The C_4 photosynthesis evolved from their C_3 ancestors with the altered expression of thousands of genes. Furthermore, many genes acquired cell-specific expression pattern. In order to develop rice with C_4 photosynthetic pathway (C_4 rice), a good understanding of the expression and cell-specific regulatory mechanism of the key C_4 genes is essential. Apart from this, the genetic information regulating the bundle-sheath (BS) cell size and BS cell chloroplasts number in C_4 plants is required in order to engineer C_4-like leaf anatomy in rice. Once all the information about genes is obtained, many genes controlling the C_4 leaf anatomy as well as the core C_4 biochemical pathway need to be redirected either in the mesophyll (M) or the BS cells of rice leaf. Considering these above-mentioned factors of the C_4-rice development project, the scientific aims of this thesis are focused mainly on the following two subprojects

(1) To build up a C_4 cycle in rice, many C_4 specific genes need to be expressed in either the M or BS cells of rice by employing cell-specific promoters. In order to discover a new set of M cell-specific promoters for rice leaf, the 5′-flanking sequences of the C_4 phospho*enol*pyruvate carboxylase (*C_4Ppc*) gene from the Panicoid grass species *Setaria viridis*, *Panicum miliaceum*, and *Urochloa maxima* were isolated through the vectorette PCR (genome-walking) technique and tested in rice. All the analyzed 5′-flanking sequences were found to be active and M cell-specific in rice leaf when fused with the β-Glucuronidase (GUS) reporter gene. Besides, different fragments of these *C_4Ppc* 5′-flanking sequences were analyzed in rice leaf tissue in order to identify the minimal promoter fragment as well as to determine the key cell-specific regulatory element(s). This was accomplished by stable transformation of the respective promoter reporter-gene constructs in rice followed by the GUS histochemical and GUS quantitative assay (Chapter 1: Manuscript 1; Gupta et al., 2018, in preparation).

(2) In an evolutionary context of C_3 to C_4 transition, anatomical preconditions such as enlarged, and organelles rich BS cell must be established before the core C_4 pathway

can evolve. Till date, only little is known about the genes regulating the BS cell ontogeny in C_4 plants. The genetic information related to photosynthetic activation of BS cells in C_4 plants must be identified in order to manipulate the rice plant for C_4 characteristics. Thus, an EMS based forward genetic screen was performed in the C_3 model plant, *Arabidopsis thaliana,* to identify genes that regulate the development of C_4 BS cells ontogeny. Prior to mutagenesis, the BS cell chloroplasts of the Arabidopsis plant were labeled with a chloroplast-targeted green fluorescent protein (GFP) using a BS cell-specific promoter (Döring, 2017). Reporter gene signal intensity was used as a proxy to identify numerous numbers of stable mutant lines with altered GFP signal. The most promising mutant lines were further analyzed microscopically to detect any alternation in the BS cell anatomy. (Chapter 2: part of the work has been used in Döring et al., 2018, submitted to *The Plant Journal*).

III. Summary

C_4 photosynthesis is a special kind of adaptation that alleviates the detrimental effects of photorespiration and has risen independently over 60 times from their C_3 ancestors. Reduction in the photorespiration reaction in C_4 photosynthetic plants is achieved through the metabolic cooperation of the mesophyll (M) and the bundle sheath (BS) cells. Crop plants that use the C_4 photosynthetic cycle are more productive and use less water and nitrogen in relation with their C_3 counterparts. These advantages of C_4 photosynthesis, along with the demand for future crop security, have spurred the development of a program to install C_4 traits into the C_3 plant rice in order to increase the latter's agricultural productivity. To achieve this goal, genes controlling the C_4 leaf anatomy and the core C_4 biochemical process need to be transferred to rice and expressed appropriately. Therefore, sets of cell-specific promoter units that can be used alternatively for the development of C_4 rice are essential. In the first part of this thesis, the 5´-flanking sequences of the C_4 phospho*enol*pyruvate carboxylase (*C_4Ppc*) gene from the Panicoid C_4 grass species *Setaria viridis*, *Panicum miliaceum* and *Urochloa maxima* have been shown to direct mesophyll cell-specific expression of a reporter gene in rice leaf tissue. 5´-deletion analysis of these flanking sequences revealed that all the essential regulatory elements responsible for directing M cell-specific expression are located within their proximal 500 bp upstream of the translational start codon ATG. Comparative sequence analysis of the *C_4Ppc* 5´-flanking sequences of six-selected C_4 grass species identified four motifs of conserved nucleotide sequences (CNS) within this proximal 500 bp. Detailed promoter deletion and recombination analysis in rice revealed that the CNSs are essential for the activity of the *C_4Ppc* promoter. Further investigation provided enough evidence that a putative interaction between the CNSs and an unknown upstream redundant element(s) directs the M cell-specific expression of the *C_4Ppc* gene in many C_4 grass species.

One of the major obstacles for developing the C_4 rice is the absence of C_4-like leaf anatomy in rice. Thereby, photosynthetic activation of BS cells must be developed before the CO_2 shuttle toward the BS cells can establish. However, very little is known about the genes controlling the BS cell size and BS chloroplast development in C_4 plants. To study them an EMS based forward genetic screen was performed in the C_3 model plant *Arabidopsis thaliana* by taking consideration of the fact that the gene

regulatory networks of the BS cells in the C_3 and C_4 dicotyledons plants are at least partly conserved. The BS cells of the Arabidopsis plants were labeled with the chloroplast-targeted green fluorescent protein (sGFP) using a BS cell-specific promoter. Differences in the reporter gene signal intensity in the Arabidopsis leaf served as a proxy for collecting mutant lines with possible altered BS cell anatomy. To this end, 10 stable EMS mutant lines were identified that showed an increased or a decreased GFP signal intensity compared to the non-mutagenized reference line. A high-resolution light microscope analysis on the six mutant lines showed that three of these mutants have an effect in the BS cell number and size of the vasculature. This study identified mutant lines with affected BS cells, and therefore it sets a good background for future research to identify genes involved in the BS cell ontogeny and maintenance.

IV. Bibliography

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., ... & Innan, H. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature biotechnology*, ***30***(2), 174.

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., & Westhoff, P. (2007). Evolution and function of a cis-regulatory module for mesophyll-specific gene expression in the C_4 dicot Flaveria trinervia. *The Plant Cell*, ***19***(11), 3391-3402.

Aliscioni, S., Bell, H.L., Besnard, G., Christin, P.A., Columbus, J.T., Duvall, M.R., Edwards, E.J., Giussani, L., Hasenstab-Lehman, K., Hilu, K.W., et al. (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytol*, ***193***: 304–312.

Andersson, I., & Taylor, T. C. (2003). Structural framework for catalysis and regulation in ribulose-1, 5-bisphosphate carboxylase/oxygenase. *Archives of Biochemistry and Biophysics*, ***414***(2), 130-140.

Aubry, S., Brown, N.J., & Hibberd, J.M. (2011). The role of proteins in C3 plants prior to their recruitment into the C_4 pathway. *Journal of experimental botany,* ***62***: 3049–3059

Bai, J., Xu, D. H., Kang, H. M., Chen, K., & Wang, G. (2008). Photoprotective function of photorespiration in Reaumuria soongorica during different levels of drought stress in natural high irradiance. *Photosynthetica,* ***46***, 232.

Bassham, A.J., Benson, A.A., & Calvin, M. (1950). The path of carbon in photosynthsis. *Journal of experimental botany*. ***185***: 781–7

Bauwe, H., Hagemann, M., Kern, R., & Timm, S. (2012). Photorespiration has a dual origin and manifold links to central metabolism. *Curr Opin Plant Biol,* ***15***: 269–275.

Bauwe, H., Hagemann, M., & Fernie, A. R. (2010). Photorespiration: players, partners and origin. *Trends in plant science*, ***15***(6), 330-336.

Bauwe, H. (2010). Chapter 6 Photorespiration: The bridge to C_4 photosynthesis. C_4 Photosynth. Relat. CO_2 Conc. Mech. *Springer Netherlands*, pp 81–108

Besnard, G., Muasya, A. M., Russier, F., Roalson, E. H., Salamin, N., & Christin, P. A. (2009). Phylogenomics of C_4 photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. *Molecular Biology and Evolution*, ***26***(8), 1909-1919.

Bläsing, O. E., Ernst, K., Streubel, M., Westhoff, P., & Svensson, P. (2002). The non-photosynthetic phosphoenolpyruvate carboxylases of the C_4 dicot Flaveria trinervia–implications for the evolution of C_4 photosynthesis. *Planta*, ***215***(3), 448-456.

Bloom, A. J., Burger, M., Rubio Asensio, J. S., & Cousins, A. B. (2010). Carbon dioxide enrichment inhibits nitrate assimilation in wheat and Arabidopsis. *Science, **328***, 899–903.

Botha, C.E. (1992). Plasmodesmatal distribution, structure and frequency in relation to assimilation in C3 and C4 grasses in southern Africa. *Planta, **187***, 348–358.

Bowes, G., Ogren, W. L., & Hageman, R. H. (1971). Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochemical and biophysical research communications*, ***45***(3), 716-722.

Bowman, J.L., Smyth, D.R. & Meyerowitz, E.M. (1991). Genetic interactions among floral homeotic genes of Arabidopsis. *Development*, ***112***, 1–20.

Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C. P., & Weber, A. P. (2014). Towards an integrative model of C4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C4 species. *Journal of experimental botany*, ***65***(13), 3579-3593.

Brodribb, T. J., & Feild, T. S. (2010). Leaf hydraulic evolution led a surge in leaf photosynthetic capacity during early angiosperm diversification. *Ecology letters*, ***13***(2), 175-183.

Brooks, A., & Farquhar, G. D. (1985). Effect of temperature on the CO 2/O 2 specificity of ribulose-1, 5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light. *Planta*, ***165***(3), 397-406.

Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., & Hibberd, J. M. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C4 photosynthesis. *Science*, ***331***(6023), 1436-1439.

Brown, W. V. (1975). Variations in anatomy, associations, and origins of Kranz tissue. *American Journal of Botany*, ***62***(4), 395-402.

Calvin, M. & Benson, A.A. (1948). The Path of Carbon in Photosynthesis. *Science*. ***107***, 476-480.

Campbell, W. J., & Ogren, W. L. (1990). Glyoxylate inhibition of ribulosebisphosphate carboxylase/oxygenase activation in intact, lysed, and reconstituted chloroplasts. *Photosynth*. *Res*. ***23***, 257–268.

Carmo-Silva, E., Scales, J. C., Madgwick, P. J., & Parry, M. A. (2015). Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant, Cell & Environment*, ***38***(9), 1817-1832.

Chollet, R., Vidal, J., & O'Leary, M. H. (1996). Phospho*enol*pyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. *Annual review of plant biology*, ***47***(1), 273-298.

Cushman, J. C., & Bohnert, H. J. (1999). Crassulacean acid metabolism: molecular

genetics. *Annual review of plant biology*, ***50***(1), 305-332.

Dalal, J., Lopez, H., Vasani, N. B., Hu, Z., Swift, J. E., Yalamanchili, R., ... & Sederoff, H. W. (2015). A photorespiratory bypass increases plant growth and seed yield in biofuel crop Camelina sativa. *Biotechnology for biofuels*, ***8***(1), 175.

Danila, F.R., Quick, W.P., White, R.G., Furbank, R.T., & von Caemmerer, S. (2016). The metabolite pathway between bundle sheath and mesophyll: quantification of plasmodesmata in leaves of C3 and C4 monocots. *The Plant Cell, **28***, 1461–1471.

Danila, F. R., Quick, W. P., White, R. G., Furbank, R. T., & von Caemmerer, S. (2018). CORRECTION: The Metabolite Pathway between Bundle Sheath and Mesophyll: Quantification of Plasmodesmata in Leaves of C3 and C4 Monocots. *The Plant Cell*, tpc-00115.

de FC Carvalho, J., Madgwick, P. J., Powers, S. J., Keys, A. J., Lea, P. J., & Parry, M. A. (2011). An engineered pathway for glyoxylate metabolism in tobacco plants aimed to avoid the release of ammonia in photorespiration. *BMC biotechnology*, ***11***(1), 111.

Dengler, N. G., & Nelson, T. (1999). Leaf structure and development in C4 plants. In '*C4 plant biology. Physiological ecology series'*. (Eds RF Sage, RK Monson) pp. 133–172.

Doebley, J., & Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. *The Plant Cell*, ***10***(7), 1075-1082.

Döring, F. (2017). *Understanding the Bundle Sheath in C4 Evolution-Forward Genetic and Transcriptomic Approaches* (Doctoral dissertation, Universitäts-und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf).

Edwards, G. E., Franceschi, V. R., & Voznesenskaya, E. V. (2004). Single-cell C4 photosynthesis versus the dual-cell (Kranz) paradigm. *Annu. Rev. Plant Biol.*, ***55***, 173-196.

Edwards, G. E., & Ku, M. S. (1987). Biochemistry of C 3–C 4 intermediates. In *Photosynthesis* (pp. 275-325).

Ehleringer, J. R., Cerling, T. E., & Helliker, B. R. (1997). C4 photosynthesis, atmospheric CO2, and climate. *Oecologia*, ***112***(3), 285-299.

Ehleringer, J. R., & Monson, R. K. (1993). Evolutionary and ecological aspects of photosynthetic pathway variation. *Annual Review of Ecology and Systematics*, ***24***(1), 411-439.

Ehleringer, J.R., Sage, R.F., Flanagan, L.B., & Pearcy, R.W. (1991). Climate change and the evolution of C4 photosynthesis. *Trends Ecol Evol, **6***: 95–9.

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., & Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. *Proc. Natl. Acad. Sci.,* ***105***, 17199–17204.

Ellis, R. J. (1979). The most abundant protein in the world. *Trends in biochemical sciences*, ***4***(11), 241-244.

Engel, N., Daele, K. van den, Kolukisaoglu, Ü., Morgenthal, K., Weckwerth, W., P.rnik, T., et al. (2007). Deletion of Glycine Decarboxylase in Arabidopsis Is Lethal under Nonphotorespiratory Conditions. *Plant Physiol*., ***144***, 1328–1335.

Engelmann, S., Zogel, C., Koczor, M., Schlue, U., Streubel, M., & Westhoff, P. (2008). Evolution of the C 4 phosphoenolpyruvate carboxylase promoter of the C 4 species Flaveria trinervia: the role of the proximal promoter region. *BMC plant biology*, ***8***(1), 4.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., ... & Westhoff, P. (2008). The gene for the P-subunit of glycine decarboxylase from the C4 species Flaveria trinervia: analysis of transcriptional control in transgenic Flaveria bidentis (C4) and Arabidopsis (C3). *Plant physiology*, ***146***(4), 1773-1785.

Ernst, K., & Westhoff, P. (1997). The phosphoenolpyruvate carboxylase (ppc) gene family of Flaveria trinervia (C 4) and F. pringlei (C 3): molecular characterization and expression analysis of the ppcB and ppcC genes. *Plant molecular biology*, ***34***(3), 427-443

Fernie, A. R., Bauwe, H., Eisenhut, M., Florian, A., Hanson, D. T., Hagemann, M., ... & Roje, S. (2013). Perspectives on plant photorespiratory metabolism. *Plant Biology*, ***15***(4), 748-753.

Furbank, R. T. (2011). Evolution of the C4 photosynthetic mechanism: are there really three C4 acid decarboxylation types? *Journal of experimental botany*, ***62***(9), 3103-3108.

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., & Westhoff, P. (2004). cis-Regulatory elements for mesophyll-specific gene expression in the C4 plant Flaveria trinervia, the promoter of the C4 phosphoenolpyruvate carboxylase gene. *The Plant Cell*, ***16***(5), 1077-1090.

Gowik, U., Engelmann, S., Bläsing, O. E., Raghavendra, A. S., & Westhoff, P. (2006). Evolution of C4 phosphoenolpyruvate carboxylase in the genus Alternanthera: gene families and the enzymatic characteristics of the C4 isozyme and its orthologues in C3 and C3/C4Alternantheras. *Planta*, ***223***(2), 359-368.

Gowik, U., & Westhoff, P. (2011). The path from C3 to C4 photosynthesis. *Plant Physiology*, ***155***(1), 56-63. d

Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C., Burtner, C., Johnson, J.E., Odden, A.R., Comai, L., and

Henikoff, S. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics, **164***: 731–740.

Griffiths, H., Weller, G., Toy, L. F., & Dennis, R. J. (2013). You're so vein: bundle sheath physiology, phylogeny and evolution in C3 and C4 plants. *Plant, Cell & Environment*, ***36***(2), 249-261.

Guan, X. Q., Zhao, S. J., Li, D. Q., and Shu, H. R. (2004). Photoprotective Function of Photorespiration in Several Grapevine Cultivars Under Drought Stress. *Photosynthetica, **42***, 31–36.

Haberlandt, G. (1882). Vergleichende anatomie des assimilatorischen gewebesystems der pflanzen. *Jb. wiss. Bot.**(13)***, 74-188.

Hagemann, M., & Bauwe, H. (2016). Photorespiration and the potential to improve photosynthesis. *Curr. Opin. Chem. Biol*., ***35***, 109–116.

Hall, L. N., Rossini, L., Cribb, L., & Langdale, J. A. (1998). GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf. *The Plant Cell*, ***10***(6), 925-936.

Hatch, M. D. (1987). C4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics*, ***895***(2), 81-106.

Hermans, J., & Westhoff, P. (1990). Analysis of expression and evolutionary relationships of phosphoenol-pyruvate carboxylase genes in Flaveria trinervia (C 4) and F. pringlei (C 3). *Molecular and General Genetics MGG*, ***224***(3), 459-468.

Hibberd, J. M., & Covshoff, S. (2010). The regulation of gene expression required for C4 photosynthesis. *Annual review of plant biology*, ***61***, 181-207.

Hylton, C. M., Rawsthorne, S., Smith, A. M., Jones, D. A., & Woolhouse, H. W. (1988). Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C3− C4 intermediate species. *Planta*, ***175***(4), 452-459.

James, G., Patel, V., Nordström, K.J., Klasen, J.R., Salomé, P.A., Weigel, D., & Schneeberger, K. (2013). User guide for mapping-by-sequencing in Arabidopsis. *Genome Biol, **14***: R61

Jordan, D. B., & Ogren, W. L. (1984). The CO2/O2 specificity of ribulose 1, 5-bisphosphate carboxylase/oxygenase. *Planta*, ***161***(4), 308-313.

Kadereit, G., Ackerly, D., & Pirie, M. D. (2012). A broader model for C4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae ss). *Proc. R. Soc. B*, ***279***(1741), 3304-3311.

Kano-Murakami, Y., Suzuki, I., Sugiyama, T., & Matsuoka, M. (1991). Sequence-specific interactions of a maize factor with a GC-rich repeat in the

phosphoenolpyruvate carboxylase gene. *Molecular and General Genetics MGG*, ***225***(2), 203-208.

Kawamura, T., Shigesada, K., Toh, H., Okumura, S., Yanagisawa, S., & Izui, K. (1992). Molecular evolution of phosphoenolpyruvate carboxylase for C4 photosynthesis in maize: comparison of its cDNA sequence with a newly isolated cDNA encoding an isozyme involved in the anaplerotic function. *The Journal of Biochemistry*, ***112***(1), 147-154.

Kebeish, R., Niessen, M., Thiruveedhi, K., Bari, R., Hirsch, H. J., Rosenkranz, R., ... & Peterhänsel, C. (2007). Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in Arabidopsis thaliana. *Nature biotechnology*, ***25***(5), 593.

Kellogg, E. A. (1999). Phylogenetic aspects of the evolution of C4 photosynthesis. *C4 plant biology*, ***12***, 411-444.

Kelly, G. J., & Latzko, E. (1976). Inhibition of spinach-leaf phosphofructokinase by 2-phosphoglycollate. *FEBS Lett*., ***68***, 55–58.

Kim, Y., Schumaker, K.S., & Zhu, J.K. (2006). EMS mutagenesis of Arabidopsis. *Methods Mol Biol, **323***: 101–103.

Kinsman, E.A., & Pyke, K.A. (1998). Bundle sheath cells and cell-specific plastid development in *Arabidopsis* leaves. Development ***125*,** 1815-1822.

Ku, S. B., & Edwards, G. E. (1977). Oxygen inhibition of photosynthesis: I. Temperature dependence and relation to O2/CO2 solubility ratio. *Plant Physiology*, *59*(5), 986-990. doi: 10.1104/pp.59.5.986

Ku, M. S., Schmitt, M. R., & Edwards, G. E. (1979). Quantitative determination of RuBP carboxylase–oxygenase protein in leaves of several C3 and C4 plants. *Journal of Experimental Botany*, ***30***(1), 89-98.

Ku, M. S., Kano-Murakami, Y., & Matsuoka, M. (1996). Evolution and expression of C4 photosynthesis genes. *Plant physiology*, ***111***(4), 949.

Lara, M. V., Chuong, S. D., Akhani, H., Andreo, C. S., & Edwards, G. E. (2006). Species having C4 single-cell-type photosynthesis in the Chenopodiaceae family evolved a photosynthetic phosphoenolpyruvate carboxylase like that of Kranz-type C4 species. *Plant Physiology*, ***142***(2), 673-684.

Langdale, J. A., Taylor, W. C., & Nelson, T. (1991). Cell-specific accumulation of maize phosphoenolpyruvate carboxylase is correlated with demethylation at a specific site> 3 kb upstream of the gene. *Molecular and General Genetics MGG*, ***225***(1), 49-55.

Latzko, E., & Kelly, G. J. (1983). The many-faceted function of phosphoenolpyruvate carboxylase in C3 plants. *Physiologie vegetale*., ***21***, 805–815

Leegood, R.C. (2008) Roles of the bundle sheath cells in leaves of C_3 plants. In *Journal of Experimental Botany*, **59,** 1663–1673.

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S. L., ... & Reidel, E. J. (2010). The developmental dynamics of the maize leaf transcriptome. *Nature genetics*, ***42***(12), 1060.

Lima Neto, M. C., Cerqueira, J. V. A., da Cunha, J. R., Ribeiro, R. V., & Silveira, J. a. G. (2017). Cyclic electron flow, NPQ and photorespiration are crucial for the establishment of young plants of Ricinus communis and Jatropha curcas exposed to drought. *Plant Biol. Stuttg. Ger.,* ***19***, 650–659.

Liu, C., Young, A. L., Starling-Windhof, A., Bracher, A., Saschenbrecker, S., Rao, B. V., ... & Beckmann, R. (2010). Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature*, ***463***(7278), 197.

Long, S. P. (1999). Environmental responses. *C4 plant biology*, 215-249.

Lundgren, M. R., Osborne, C. P., & Christin, P. A. (2014). Deconstructing Kranz anatomy to understand C4 evolution. *Journal of experimental botany*, ***65***(13), 3357-3369.

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, ***290***(5494), 1151-1155.

Majeran, W., Friso, G., Ponnala, L., Connolly, B., Huang, M., Reidel, E., ... & Turgeon, R. (2010). Structural and metabolic transitions of C4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. *The Plant Cell*, ***22***(11), 3509-3542.

Marcus, Y., Altman-Gueta, H., Wolff, Y., & Gurevitz, M. (2011). Rubisco mutagenesis provides new insight into limitations on photosynthesis and growth in Synechocystis PCC6803. *Journal of experimental botany*, ***62***(12), 4173-4182.

Masumoto, C., Miyazawa, S. I., Ohkawa, H., Fukuda, T., Taniguchi, Y., Murayama, S., ... & Miyao, M. (2010). Phosphoenolpyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. *Proceedings of the National Academy of Sciences*, ***107***(11), 5226-5231.

Matsuoka, M., Kyozuka, J., Shimamoto, K., & Kano-Murakami, Y. (1994). The promoters of two carboxylases in a C4 plant (maize) direct cell-specific, light-regulated expression in a C3 plant (rice). *The Plant Journal*, ***6***(3), 311-319.

Matsuoka, M., Tada, Y., Fujimura, T., & Kano-Murakami, Y. (1993). Tissue-specific light-regulated expression directed by the promoter of a C4 gene, maize pyruvate, orthophosphate dikinase, in a C3 plant, rice. *Proceedings of the National Academy of Sciences*, ***90***(20), 9586-9590.

McKown, A. D., Moncalvo, J. M., & Dengler, N. G. (2005). Phylogeny of Flaveria

(Asteraceae) and inference of C4 photosynthesis evolution. *American Journal of Botany*, ***92***(11), 1911-1928.

Melzer, E., & O'Leary, M. H. (1987). Anapleurotic CO2 fixation by phosphoenolpyruvate carboxylase in C3 plants. *Plant Physiology*, ***84***(1), 58-60.

Mitchell, P. L., & Sheehy, J. E. (2006). Supercharging rice photosynthesis to increase yield. *New Phytologist*, *171*(4), 688-693. doi: 10.1111/j.1469-8137.2006.01856.x

Miyao, M., Masumoto, C., Miyazawa, S. I., & Fukayama, H. (2011). Lessons from engineering a single-cell C4 photosynthetic pathway into rice. *Journal of experimental botany*, *62*(9), 3021-3029. doi: 10.1093/jxb/err023

Monson, R. K., Edwards, G. E., & Ku, M. S. (1984). C3-C4 intermediate photosynthesis in plants. *Bioscience*, *34*(9), 563-574. doi: 10.2307/1309599

Muhaidat, R., Sage, T. L., Frohlich, M. W., Dengler, N. G., & Sage, R. F. (2011). Characterization of C3–C4 intermediate species in the genus Heliotropium L.(Boraginaceae): anatomy, ultrastructure and enzyme activity. *Plant, Cell & Environment*, ***34***(10), 1723-1736.

Newman, J., Brändén, C. I., & Jones, T. A. (1993). Structure determination and refinement of ribulose 1, 5-bisphosphate carboxylase/oxygenase from Synechococcus PCC6301. *Acta Crystallographica Section D: Biological Crystallography*, ***49***(6), 548-560.

Norman, E. G., and Colman, B. (1991). Purification and Characterization of Phosphoglycolate Phosphatase from the Cyanobacterium Coccochloris peniocystis. *Plant Physiol*. 95, 693– 698. doi:10.1104/pp.95.3.693.

Novitskaya, L., Trevanion, S. J., Driscoll, S., Foyer, C. H., & Noctor, G. (2002). How does photorespiration modulate leaf amino acid contents? A dual approach through modelling and metabolite analysis. *Plant, Cell & Environment*, *25*(7), 821-835.

Oaks, A. (1994). Efficiency of nitrogen utilization in C3 and C4 cereals. *Plant Physiology*, ***106***(2), 407.

Parry, M. A. J., Andralojc, P. J., Mitchell, R. A., Madgwick, P. J., & Keys, A. J. (2003). Manipulation of Rubisco: the amount, activity, function and regulation. *Journal of experimental botany*, *54*(386), 1321-1333. doi: 10.1093/jxb/erg141.

Parry, M. A., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H., & Whitney, S. M. (2012). Rubisco activity and regulation as targets for crop improvement. *Journal of Experimental Botany*, *64*(3), 717-730.

Peterhansel, C., Horst, I., Niessen, M., Blume, C., Kebeish, R., Kürkcüoglu, S., & Kreuzaler, F. (2010). Photorespiration. *The Arabidopsis Book*, e0130.

Pick, T. R., Br.utigam, A., Schulz, M. A., Obata, T., Fernie, A. R., and Weber, A. P. M. (2013). PLGG1, a plastidic glycolate glycerate transporter, is required for

photorespiration and defines a unique class of metabolite transporters. *Proc. Natl. Acad. Sci. U. S. A*. ***110***, 3185–3190.

Rachmilevitch, S., Cousins, A. B., & Bloom, A. J. (2004). Nitrate assimilation in plant shoots depends on photorespiration. *Proceedings of the National Academy of Sciences of the United States of America*, ***101***(31), 11506-11510.

Rademacher, N., Kern, R., Fujiwara, T., Mettler-Altmann, T., Miyagishima, S., Hagemann, M., et al. (2016). Photorespiratory glycolate oxidase is essential for the survival of the red alga Cyanidioschyzon merolae under ambient CO_2 conditions. *J. Exp. Bot.* 67, 3165–3175.

Raines CA. 2011. Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. *Plant Physiology* 155, 36–42

Rao, X., & Dixon, R. A. (2016). The differences between NAD-ME and NADP-ME subtypes of C4 photosynthesis: more than decarboxylating enzymes. *Frontiers in plant science*, ***7*****,** 1525.

Roalson, E. H., Hinchliff, C. E., Trevisan, R., & da Silva, C. R. (2010). Phylogenetic relationships in Eleocharis (Cyperaceae): C4 photosynthesis origins and patterns of diversification in the spikerushes. *Systematic Botany*, *35*(2), 257-271.

Sage, R. F., & Pearcy, R. W. (1987). The nitrogen use efficiency of C3 and C4 plants: II. Leaf nitrogen effects on the gas exchange characteristics of Chenopodium album (L.) and Amaranthus retroflexus (L.). *Plant Physiology*, ***84***(3), 959-963.

Sage, R. F. (2004). The evolution of C4 photosynthesis. *New phytologist*, ***161***(2), 341-370.

Sage, R. F., Christin, P. A., & Edwards, E. J. (2011). The C4 plant lineages of planet Earth. *Journal of Experimental Botany*, ***62***(9), 3155-3169.

Sage, R. F., Sage, T. L., & Kocacinar, F. (2012). Photorespiration and the evolution of C4 photosynthesis. *Annual review of plant biology*, ***63***, 19-47.

Sage, R. F. (2016). A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of experimental botany*, ***68***(2), e11-e28.

Schaffner, A. R., & Sheen, J. (1992). Maize C4 photosynthesis involves differential regulation of phosphoenolpyruvate carboxylase genes. *The Plant Journal*, ***2***(2), 221-232.

Schuler, M. L., Mantegazza, O., & Weber, A. P. (2016). Engineering C4 photosynthesis into C3 chassis in the synthetic biology age. *The Plant Journal*, ***87***(1), 51-65.

Schuller, K. A., Plaxton, W. C., & Turpin, D. H. (1990). Regulation of phosphoenolpyruvate carboxylase from the green alga Selenastrum minutum:

properties associated with replenishment of tricarboxylic acid cycle intermediates during ammonium assimilation. *Plant Physiology*, ***93***(4), 1303-1311.

Scoffoni, C., Rawls, M., McKown, A., Cochard, H., & Sack, L. (2011). Decline of leaf hydraulic conductance with dehydration: relationship to leaf size and venation architecture. *Plant Physiology*, ***156***(2), 832-843.

Sharkey, T. D. (1988). Estimating the rate of photorespiration in leaves. *Physiol. Plant*. 73, 147–152.

Sheen, J. (1999). C4 gene expression. *Annual review of plant biology*, ***50***(1), 187-217.

Sheehy, J. E., Ferrer, A. B., & Mitchell, P. L. (2008). Harnessing photosynthesis in tomorrow's world: humans, crop production and poverty alleviation. In *Photosynthesis. Energy from the sun* (pp. 1237-1242). Springer, Dordrecht.

Sikora P, Chawade A, Larsson M, Olsson J, Olsson O, Sikora P, Chawade A, Larsson M, Olsson J, Olsson O (2011). Mutagenesis as a tool in plant genetics, functional genomics, and breeding. *Int J Plant Genomics, **2011***: 314829

Slewinski, T. L. (2013). Using evolution as a guide to engineer Kranz-type C4 photosynthesis. *Frontiers in plant science*, ***4***, 212.

Somerville, C.R. (2001). An Early Arabidopsis Demonstration. Resolving a Few Issues Concerning Photorespiration. *Plant Physiology, **125***: 20-24

Sparks, E. E., Drapek, C., Gaudinier, A., Li, S., Ansariola, M., Shen, N., ... & Foret, J. (2016). Establishment of expression in the SHORTROOT-SCARECROW transcriptional Cascade through opposing activities of both activators and repressors. *Developmental cell*, ***39***(5), 585-596.

Stockhaus, J., Schlue, U., Koczor, M., Chitty, J. A., Taylor, W. C., & Westhoff, P. (1997). The Promoter of the Gene Encoding the C4 Form of Phosphoenolpyruvate Carboxylase Directs Mesophyll-Specific Expression in Transgenic C4 Flaveria spp. *The Plant Cell*, ***9***(4), 479-489.

Svensson, P., Bläsing, O. E., & Westhoff, P. (2003). Evolution of C4 phosphoenolpyruvate carboxylase. *Archives of Biochemistry and Biophysics*, ***414***(2), 180-188.

Taniguchi, M., Izawa, K., Ku, M. S., Lin, J. H., Saito, H., Ishida, Y., ... & Sugiyama, T. (2000). Binding of cell type-specific nuclear proteins to the 5′-flanking region of maize C4 phospho*enol*pyruvate carboxylase gene confers its differential transcription in mesophyll cells. *Plant molecular biology*, ***44***(4), 543-557.

Taniguchi, Y., Ohkawa, H., Masumoto, C., Fukuda, T., Tamai, T., Lee, K., ... & Miyao, M. (2008). Overproduction of C4 photosynthetic enzymes in transgenic rice plants: an approach to introduce the C4-like photosynthetic pathway into rice. *Journal of experimental botany*, ***59***(7), 1799-1809.

Till, B.J., Cooper, J., Tai, T.H., Colowit, P., Greene, E.A., Henikoff, S., and Comai, L. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol*., 7: 19.

Ting, I. P., & Osmond, C. B. (1973). Multiple forms of plant phosphoenolpyruvate carboxylase associated with different metabolic pathways. *Plant physiology*, ***51***(3), 448-453.

Ueno, O., Kawano, Y., Wakayama, M., & Takeda, T. (2006). Leaf vascular systems in C3 and C4 grasses: a two-dimensional analysis. *Annals of Botany*, ***97***(4), 611-621.

United Nations. (2015). World Population Prospects The 2015 Revision.

von Caemmerer, S., & Evans, J. R. (2010). Enhancing C3 photosynthesis. *Plant Physiology*, ***154***(2), 589-592.

von Caemmerer, S., Edwards, G. E., Koteyeva, N., & Cousins, A. B. (2014). Single cell C 4 photosynthesis in aquatic and terrestrial plants: A gas exchange perspective. *Aquatic botany*, ***118***, 71-80

Walker, B. J., VanLoocke, A., Bernacchi, C. J., and Ort, D. R. (2016). The Costs of Photorespiration to Food Production Now and in the Future. Annu. Rev. Plant Biol., ***67***, 107–129.

Wang, Y., Bräutigam, A., Weber, A. P., & Zhu, X. G. (2014). Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. *Journal of experimental botany*, ***65***(13), 3567-3578.

Wang, P., Khoshravesh, R., Karki, S., Tapia, R., Balahadia, C. P., Bandyopadhyay, A., ... & Langdale, J. A. (2017). Re-creation of a Key Step in the Evolutionary Switch from C3 to C4 Leaf Anatomy. *Current Biology*, ***27***(21), 3278-3287.

Waters, M. T., Moylan, E. C., & Langdale, J. A. (2008). GLK transcription factors regulate chloroplast development in a cell-autonomous manner. *The Plant Journal*, ***56***(3), 432-444.

Waters, M. T., & Langdale, J. A. (2009). The making of a chloroplast. *The EMBO journal*, ***28***(19), 2861-2873.

Westhoff, P., & Gowik, U. (2010). Evolution of C4 photosynthesis—looking for the master switch. *Plant Physiology*, ***154***(2), 598-601.

Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., ... & Westhoff, P. (2012). Regulation of the photorespiratory GLDPA gene in C4 Flaveria: an intricate interplay of transcriptional and posttranscriptional processes. *The Plant Cell*, ***24***(1), 137-151.

Windhövel, A., Hein, I., Dabrowa, R., & Stockhaus, J. (2001). Characterization of a novel class of plant homeodomain proteins that bind to the C4 phosphoenolpyruvate carboxylase gene of Flaveria trinervia. *Plant Molecular Biology*, ***45***(2), 201-214.

Winter, K. (1985). Crassulacean acid metabolism. *Photosynthetic Mechanisms and the Environment.*, 329-387.

Yanagisawa, S., & Sheen, J. (1998). Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. *The Plant Cell*, ***10***(1), 75-89.

Yoshimura, Y., Kubota, F., & Ueno, O. (2004). Structural and biochemical bases of photorespiration in C4 plants: quantification of organelles and glycine decarboxylase. *Planta*, ***220***(2), 307-317.

Zelitch, I., Schultes, N. P., Peterson, R. B., Brown, P., & Brutnell, T. P. (2009). High glycolate oxidase activity is required for survival of maize in normal air. *Plant Physiology*, ***149***(1), 195-204.

Zhu, X. G., Long, S. P., & Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annual review of plant biology*, ***61***, 235-261.

V. Chapters

1. **Manuscript 1: Shipan Das Gupta, Myles Levey, Stefanie Schulze, Shanta Karki, Jan Emmerling, Monika Streubel, Paul Quick & Peter Westhoff (2018)**. The C_4 Phospho*enol*pyruvate carboxylase (*C_4Ppc*) promoters of many C_4 grass species share a common regulatory mechanism to express gene in mesophyll cell. Publication in preparation for *The Plant journal*.

2. "Screening for ethylmethane sulfonate (EMS) mutants of *Arabidopsis thaliana* affected in the anatomy of bundle sheath cells". Part of this work has been used in the-

   Florian Döring, Kumari Billakurthi, Udo Gowik, Stefanie Sultmanis, Roxana Khoshravesh, Shipan Das Gupta, Tammy Sage & Peter Westhoff (2018). Reporter based genetic screen to identify bundle sheath anatomy mutants in *Arabidopsis thaliana.* Submitted to *The Plant Journal*

Chapter 1

Manuscript 1

The C_4 Phospho*enol*pyruvate carboxylase (*C_4Ppc*) promoters of many C_4 grass species share a common regulatory mechanism to express gene in the mesophyll cell

Running title: The *C4Ppc* promoters of many C4 grass species share a common regulatory mechanism for gene expression in the mesophyll cell

Shipan Das Gupta[1], Myles Levey[1], Stefanie Schulze[1], Shanta Karki[2,3], Jan Emmerling[1], Monika Streubel[1], Udo Gowik[1,4], Paul Quick[2] & Peter Westhoff[1*]

[1]Institute of Plant Molecular and Developmental Biology, Universitätsstrasse 1, Heinrich-Heine-University, D-40225 Düsseldorf, Germany.

[2] International Rice Research Institute, 4031 Los Banos, Laguna, Philippines.

[3] National Citrus Development Program, Kirtipur, Kathmandu, Nepal

[4] Department of Biology and Environmental Sciences, Carl Von Ossietzky University, Ammerländer Heerstrasse 114, D-26129 Oldenburg, Germany

*To whom should be addressed

Shipan Das Gupta
e-mail: shipan.gupta@hhu.de

Myles Levey
e-mail: myles.levey@hhu.de

Stefanie Schulze
e-mail: stefanie.schulze@hhu.de

Shanta Karki
e-mail: shantakyoto@gmail.com

Jan Emmerling
e-mail: jan.emmerling@hhu.de

Monika Streubel
e-mail: monika.Streubel@hhu.de

Udo Gowik
udo.gowik@uni-oldenburg.de

Paul Quick
e-mail: w.p.quick@irri.org

Peter Westhoff
e-mail: west@hhu.de

Publication in preparation for *The Plant Journal*

(i) Abstract

C_4 photosynthetic plants have better radiation, nitrogen and water-use efficiencies compared to C_3 plants. Installing a C_4 pathway into C_3 crops, like rice, is assumed to increase their yield by 50 %. To establish a C_4 photosynthesis engine in rice, many genes that regulate C_4 leaf anatomy and biochemical processes have to be transformed in rice. Therefore, expressions of many genes have to be redirected either in mesophyll (M) cells or in bundle sheath (BS) cells and expressed in an appropriate manner. Thereby, it seems inevitable to discover a set of cell-specific promoters, which can be utilized alternatively for the development of a multi-transgenic C_4 plant. In this study we demonstrated that, the 5′-flanking sequences of the C_4 type phospho*enol*pyruvate carboxylase (*Ppc*) gene from three C_4 grass species can drive M cell-specific expression of a reporter gene in rice. In addition to that, we identified the minimal region of the analyzed promoters that contains all the essential regulatory elements for driving M cell-specific expression. Moreover, four motifs of conserved nucleotide sequences (CNS)s were also determined, which are essential for the activity of promoter. A putative interaction between the CNSs and an unknown redundant upstream element(s) is required for driving M cell-specific expression. This work identifies not only the three new promoters for the C_4 rice project, but also determine the unknown mystery of the *C_4Ppc* regulatory mechanism of many C_4 grass species.

Keywords: C_4 photosynthesis, C_4 rice, mesophyll (M) cell, phospho*enol*pyruvate carboxylase (*Ppc*) gene, conserved nucleotide sequence (CNS).

(ii) Introduction

More than 7.5 billion people are currently living on earth and the number is increasing day by day. It has been estimated that world population will reach almost 10 billion by 2050 (United Nations, 2015). The rising population means a higher demand of food, land and water. Already poverty and hunger have become big issues worldwide and more than 815 million people are suffering from chronic hunger every day (FAO IFAD UNICEF, 2017). Beside this, decreasing farmland areas due to industrialization and unpredictable climate changes are constantly challenging on agricultural food production. With the increasing number of people in this world, it is highly demanded to produce more crop yield in the near future.

Rice is a crop plant majorly grown in Asia. The inhabitants of Asia largely depend on rice for their caloric intake. In the near future, Asia must increase rice yield production by 60 % to provide adequate nutrition to its growing population (Sheehy et al., 2008) . However, the current breeding approaches will likely be insufficient to increase rice production by the 60 % required in the year of 2050 (Sheehy, 2001). Thus, a second green revolution is required to accelerate the increase in production of rice yield. Since the two-celled C_4 photosynthetic plants produce 50 % more yield in comparison to C_3 plants (Hibberd et al., 2008; Mitchell & Sheehy, 2006; Sage, 2004) engineering of the C_4 pathway into C_3 crop plants like rice could be a solution to increase crop productivity (Kajala et al., 2011; Karki et al., 2013; Schuler et al., 2016). For this purpose, the "International C_4 Rice Consortium" was founded in 2008 led by the International Rice Research Institute (IRRI) with 24 participating research groups. The main aim of this "C_4 Rice Consortium" is to test the feasibility of installing a NADP-ME type C_4 photosynthetic biochemical pathway into C_3 rice (http://irri.org/c4rice). There is no doubt that installing the C_4 engine into C_3 rice is an enormous challenge. To establish a C_4 metabolism in rice, a set of genes which regulate C_4 leaf anatomy and biochemical processes have to be transformed into rice and expressed in an appropriate manner. For this, expression of several genes has to be redirected to either mesophyll (M) or bundle sheath (BS) cells. Therefore, a set of cell-specific promoter units is highly needed that can be used alternatively for the generation of C_4 rice. Only two promoters (promoter of C_4-type phospho*enol*pyruvate carboxylase and phospho*enol*pyruvate di-kinase genes from maize) are reported and extensively used so far, which confer M cell-specificity in the rice leaf tissue at a desired level (Matsuoka et al., 1994, 1993). In contrast, the *Zoysia japonica* phospho*enol*pyruvate carboxykinase (*PCK*) promoter is the only reported promoter that produces BS-specific expression in rice (Nomura et al., 2005). Though, it is still not known how many genes are needed to be expressed in a cell-specific manner in rice to implement the C_4 pathway. It was proposed that at least 14 genes, including the core components of the C_4 cycle and key transporters are required (Covshoff & Hibberd, 2012; Weber & von Caemmerer, 2010) for a functioning C_4 cycle in C_3 plants. However, comparing the transcriptomes of the C_3 (*Cleome spinosa*) and the C_4 (*Cleome gyandra*) plants showed that, up to 603 transcripts are differed in abundance between the C_3 and C_4 leaves (Brautigam et al., 2011). In another study, the transcriptomics data of the *Flaveria* genus showed, 3583 transcripts are expressed differentially in C_3 and C_4 species

(Gowik et al., 2011). This indicates the numbers of genes required for the C_4 rice development are probably in the range of 14 to 3583. But the current toolbox available to drive cell-specific expression of foreign genes in rice might not be sufficient for this huge project. Using a same promoter several times for the development of multi-transgenic C_4 rice may cause silencing effects by repetitive elements (Assaad et al., 1993; Hsieh & Fire, 2000). Thereby, alternative M and BS cell-specific promoters are needed for the sake of C_4 rice development. The 5´-flanking sequences of M and BS cell-specific genes of different C_4 plants could be valuable resources for discovering new cell-specific promoters for rice.

Phospho*enol*pyruvate carboxylase (PEPC; EC.4.1.1.31) is an important enzyme in C_4 photosynthesis. It catalyzes the initial fixation of CO_2 into oxaloacetate, a C_4 compound. In plants four different isozymes of PEPC have been recognized; a C_4 type, a C_3 type, a root type and a bacterial type (Ku et al., 1996). The C_4 type PEPC serves the initial fixation of CO_2 in the C_4 pathway and is exclusively expressed in the M cells of C_4 leaves (Ku et al., 1996). The photosynthetic C_4 isoform of *Ppc* gene has evolved independently several times from the ancestral C_3 non-photosynthetic isoforms (Kellogg EA, 1999). Non-photosynthetic isoforms of *Ppc* with different catalytic and metabolic activities are found in both photosynthetic and non-photosynthetic tissue of all plants (Cushman & Bohnert, 1999; Latzko & Kelly, 1983; Melzer & O'Leary, 1987; Schuller et al., 1990). In comparison to the high expression of *C_4Ppc* in the M cells, the non-photosynthetic *Ppc* isoforms are expressed moderately in all plant tissues (Ernst & Westhoff, 1997; Hermans & Westhoff, 1990; Kawamura, 1992). Although there are major differences in the cell-specificity of C_3 and C_4 type *Ppc*s', the structure of the coding sequences is highly conserved between the C_3 and C_4 type *Ppc* (Chollet et al., 1996; Gowik & Westhoff, 2011). Strong M cell-specific expression of the *C_4Ppc* must have occurred during the evolution from the ancestral C_3 isoforms. There is compelling evidence that cell-specific expression of the *C_4Ppc* might have evolved by changes in the transcriptional control of C_3 isoforms (Sheen, 1999; Stockhaus et al., 1997). Studies on the genus *Flaveria* revealed that, the M cell-specific expression of the C_4 *ppc*A gene is indeed determined by a 41 bp segment located in the distal region of the promoter, known as mesophyll expression module-1 (MEM-1). A similar MEM-1 homologue can also be found in the *ppc*A promoter of the C_3 *Flaveria* species. However, the C_3 and C_4 specific MEM-1 differ at two positions, a G to A transition and an insertion of the

tetranucleotide CACT in the C_4 MEM-1 in comparison to the C_3 one. These two changes in the MEM-1 sequence are mandatory for the evolution of M cell-specific C_4 *ppc*A in the genus *Flaveria* (Akyildiz et al., 2007; Gowik et al., 2004). In another study it was reported that the *Zea mays C4Ppc* promoter could drive strong M cell-specific expression of a reporter gene in the C_3 rice plant (Matsuoka et al., 1994). This demonstrated that M cell-specific expression of the *C4Ppc* gene of *Z. mays* is mainly determined on the transcription level. Comparison of the C_3 and C_4 type 5′-flanking sequences of *Z. mays Ppc*s' revealed that the sequences are identical at a range from the TATA box element to the translational start codon ATG (Schäffner & Sheen, 1992). Diverse sequences upstream of their TATA box indicate that the *C4Ppc* gene of *Z. mays* might have evolved from an ancestral *C3Ppc* gene after an unequal recombination event near the TATA box element (Schäffner & Sheen, 1992; Sheen, 1999). This recombination event might have brought all the *cis* elements necessary for the M cell-specific expression of *C4Ppc*. Information available on the eudicot *Flaveria C4Ppc*A and the grass *Z. mays C4Ppc* suggests that the M cell-specific expression of *C4Ppc* genes is largely controlled by the promoter sequences. Thereby, the *C4Ppc* 5′-flanking sequences from different C_4 species could be valuable resources for discovering alternative M cell-specific drivers for rice.

Previously, it was reported that the *C4Ppc* (*ZmPpc*) and the pyruvate ortho-phosphate di-kinase (*PPDK*) (*ZmPPDK*) promoters of C_4 grass species *Zea mays* generate strong M cell-specific expression in rice (Matsuoka et al., 1994, 1993). One can assume that, the *C4Ppc* promoters from phylogenetically closely related C_4 species of *Z. mays* might also be able to drive similar M cell-specific expression in rice. Those species could be found in the Panicoideae family. The Panicoideae are one of the subfamilies of the PACMAD clade and contains diverse sets of C_4 grass species (Christin et al., 2009). Most of the commercially important C_4 crop plants like *Zea mays*, *Sorghum officinare* etc. belong to the Panicoideae. In our present study, we investigated the 5′-flanking sequences of the *C4Ppc* genes from the Panicoideae C_4 grass species *Urochloa maxima*, *Panicum miliaceum* and *Setaria viridis*. The three 5′-flanking sequences showed M cell-specific activity in rice when fused with the *uid*A (GUS) reporter gene. To fully understand the anatomy of these isolated *C4Ppc* 5′-flanking sequences, detailed promoter reporter gene analysis was performed in rice. The experiments revealed that, the proximal 500 bp of these *C4Ppc* 5′-flanking

sequences are sufficient to drive M cell-specific expression of the reporter gene in rice. Further analysis of the proximal 500 bp identified four motifs of conserved nucleotide sequence (CNS). Promoter deletion and recombination analysis in rice revealed that, these four CNSs are very important for the *C4Ppc* promoter activity. Presence of the CNSs in the six-analysed Panicoid C_4 grass species indicates a high degree of conservation of the *C4Ppc* regulatory mechanism. Identification of the CNSs is, therefore, the first characterized putative regulatory modules for *C4Ppc* genes from grass plants. In this article we describe, how we discovered a set of alternative promoters that direct M cell-specific expression in rice. We also address what the putative core regulatory element(s) of these promoters responsible for directing the activity and M cell-specificity are.

(iii) Results

Panicoideae C_4 grass C_4*Ppc* 5′-flanking sequences are active and M cell-specific in rice

In order to discover alternative sets of M cell-specific promoters for rice, the 5′-flanking sequences of the *C4Ppc* gene were isolated through vectorette PCR (genome walking technique) from the Panicoid C_4 grass species *Setaria viridis* (*SvPpc*), *Panicum miliaceum* (*PmPpc*) and *Urochloa maxima* (*UmPpc*), which exhibit NADP-ME, NAD-ME and PCK types of C_4 photosynthesis respectively (Siebert et al., 1995). The length of the three isolated 5′-flanking sequences of *SvPpc*, *PmPpc* and *UmPpc* were 2233 bp, 3327 bp and 3531 bp respectively (Figure 1A). To test the activity of these 5′-flanking sequences in rice, the three putative promoters were fused to the coding region of the *uid*A (GUS) gene linked to the terminator of the nopaline-synthase (Nos) gene (Figure 1A). In addition, *ZmPpc* (1283 bp) was used as a control construct for the M cell-specific GUS expression in rice. All the chimeric constructs were stably transformed in the *Oryza sativa ssp. japonica* Nipponbare cultivar for analysis. Like the *ZmPpc* promoter these three *C4Ppc* 5′-flanking sequences (*SvPpc*, *PmPpc* and *UmPpc*) show expression of the reporter gene in the rice M cells. GUS histochemical analysis of these three putative *C4Ppc* promoters showed that they are able to keep their C_4 characteristic M cell-specific expression pattern in rice (Figure 1B). No blue GUS staining was observed in upper or lower epidermis, vascular tissue and BS cells of the rice leaf blade

cross sections (Figure 1B). Although the GUS expression pattern of the three putative promoters in rice is very similar, the level of GUS activity driven by the

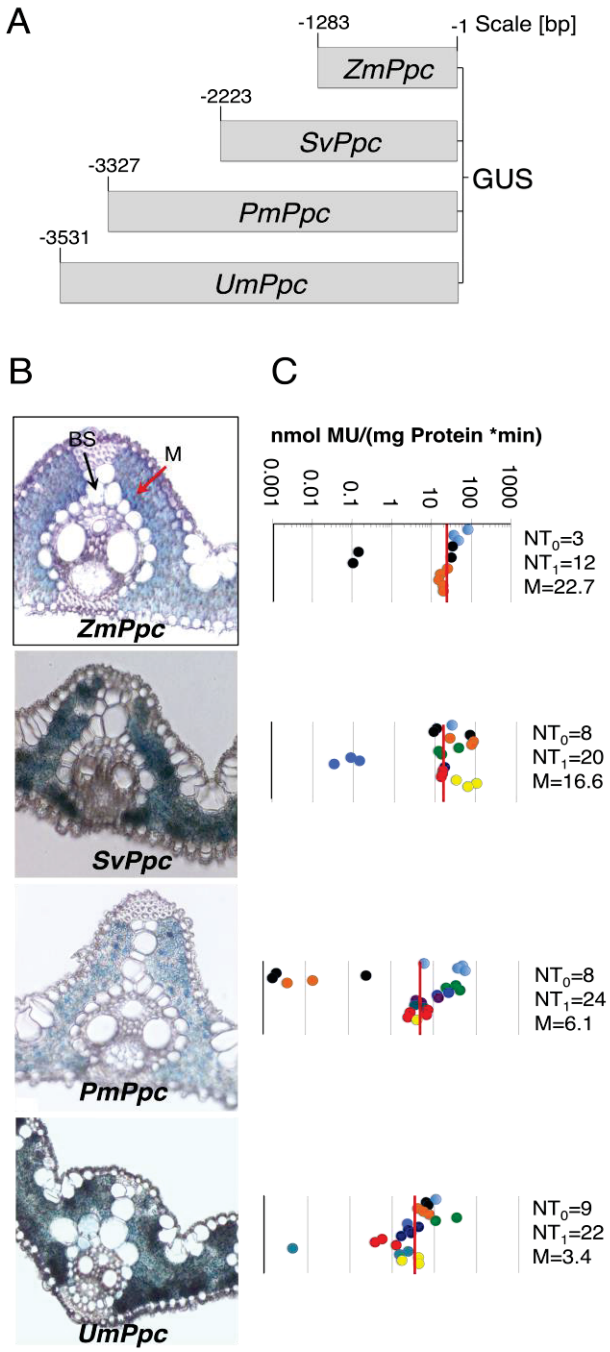


Figure 1: Analysis of the *C4Ppc* 5′-flanking sequences from *Z. mays*, *S. viridis*, *P. miliaceum* and *U. maxima* in rice reveals M cell-specific expression.

(A) Schematic presentation of the *C4Ppc* 5′-flanking sequences GUS reporter gene constructs from Panicoideae species *Z. mays*, *S. virdis*, *P. miliaceum* and *U. maxima.* Nucleotide numbers refer to the translational start codon (ATG). **(B)** Histochemical

analysis of GUS in leaf cross sections of transgenic rice plants transformed with *ZmPpc*:GUS, *SvPpc*:GUS, *PmPpc*:GUS and *UmPpc*:GUS constructs respectively. Black arrow points at the M cells of rice leaf blades and the red arrow shows the BS cells. **(C)** Quantitative measurement of GUS activities in leaves of transgenic rice plants. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. Each single dot represents one T_1 transgenic plant and the same color in dots indicates the plants belong to the same transgenic line. The total number of transgenic plants investigated (NT_1) from the independent transgenic (NT_0) lines and the median value (M) of GUS activity are represented at the right side of each column. Relative positions of the median values in each column are marked in black lines.

three putative promoters are slightly different from each other. In comparison to the *ZmPpc* promoter, the *SvPpc* 5´-flanking sequences drove a comparable type of GUS activity in rice leaf blades. In contrast, the *PmPpc* and the *UmPpc* 5´-flanking sequences showed 3.5- and 7-fold lower GUS activities respectively compared to the activity of the *ZmPpc* promoter (Figure 1C). Nevertheless, the GUS histochemical and quantitative analysis data suggest that the *trans*-regulatory systems responsible to direct M cell-specific expressions of *C_4Ppc* genes in C_4 grass plants are also present in the C_3 plant rice. Therefore, the isolated 5´-flanking sequences can be used as M cell-specific drivers for rice leaf tissue.

The proximal 500 bp of the 5´-flanking sequences of *UmPpc* are capable to drive M cell-specificity in rice leaf

The three putative *C_4Ppc* promoters are relatively huge in size. Since the engineering of C_4 photosynthesis in a rice plant might require the transformation of an array of genes the size of promoters should be minimal in order to reduce the complexity of transformation methods. The transformation efficiency and accuracy by the current gene transformation method in rice are largely depending on the construct size (Chan et al., 2002) . Thus, utilizing the minimal size of promoters would be useful to restrict the construct size in a considerable range. Therefore, the core regulatory element(s) responsible for the activity and M cells-specificity of these three putative promoters (5´-flanking sequences) are to be resolved. In order to understand the anatomy of a grass *C_4Ppc* promoter we used the *UmPpc* promoter as a starting point

and aimed to identify the minimal set of *cis* regulatory element(s) that are essential for its M cell-specificity. To rule out that, a detailed promoter reporter gene analysis was performed in transgenic C_3 rice. The full-length 5´-flanking sequences of *UmPpc* was 3531 bp long. To delimit it, the 5´-flanking sequences of *UmPpc* was shortened into three fragments with the length of 1500 bp, 1000 bp, and 500 bp respectively and fused to the GUS reporter gene. The resulting chimeric constructs *UmPpc*-1500:GUS, *UmPpc*-1000:GUS and *UmPpc*-500:GUS were transformed in the C_3 *Oryza sativa ssp. japonica* Kitaake cultivar for analysis.

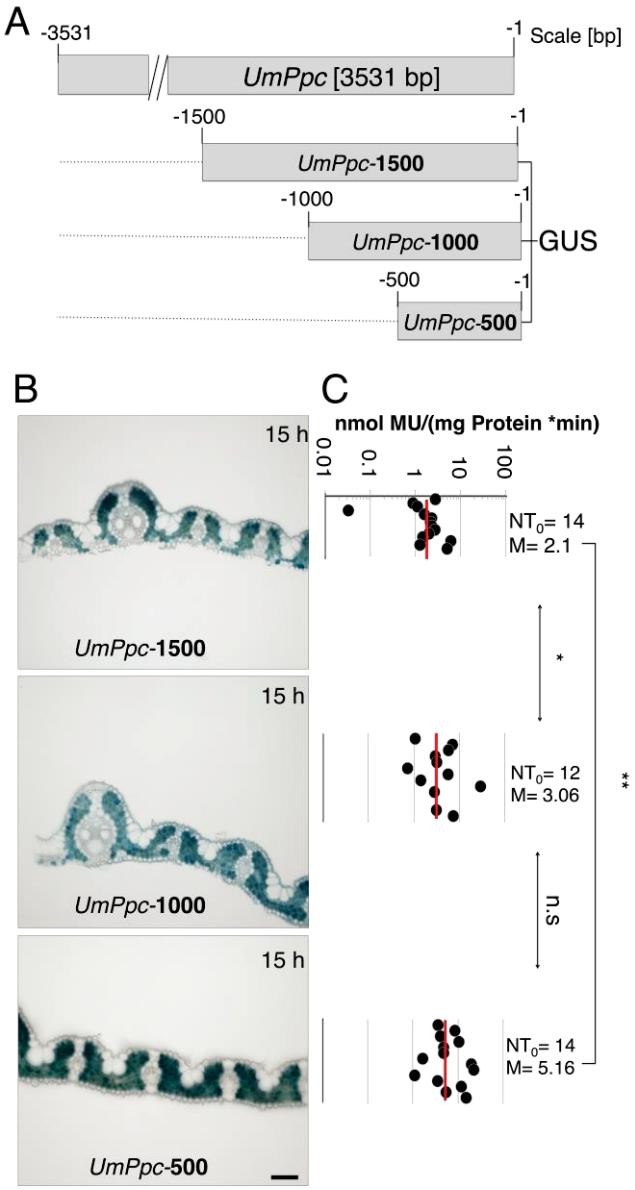


Figure 2: Deletion analysis of the *UmPpc* 5´-flanking sequences for promoter activity in transgenic rice leaves.

(A) Schematic presentation of the three chimeric constructs of the *UmPpc* 5´-flanking sequences. Dissected *UmPpc* 5'-flanking sequences with the length of 1500 bp, 1000 bp and 500 bp respectively were fused upstream of the reporter gene GUS. **(B)** Histochemical localization of GUS (blue staining) in leaf cross sections of transgenic rice transformed with *UmPpc*-1500:GUS, *UmPpc*-1000:GUS and *UmPpc*-500:GUS. Incubation period was 15 hours. Scale bar: 50 µm. **(C)** Fluorometrical quantification of GUS activity of the truncated *UmPpc* (1500 bp, 1000 bp and 500 bp) promoter fragments. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. Each single dot represents one independent transgenic plant. The number of independent transgenic plants analyzed (NT_0) and the median value (M) of GUS activity are represented at the right side of each column. The red lines in the graph indicate the median value of GUS activity. (* $p < 0.05$; ** $p < 0.005$; n.s., not significant $p > 0.05$).

Mature leaves from the transgenic T_0 rice plants were used for GUS activity analysis. Histochemical staining of leaf cross-sections of the *UmPpc*-1500:GUS, *UmPpc*-1000:GUS and *UmPpc*-500:GUS transformants showed GUS activity only in the M cells (Figure 2B). The same staining pattern was previously observed in the transformants of full-length *UmPpc* 5´-flanking sequences (Figure 1B). Fluorometrical quantification of GUS activity of the three constructs lead to almost equal amounts of GUS activity (Figure 2C). There is no notable difference among the three dissected promoters in terms of overall activity and M cell-specificity. We concluded from these results that the *cis* regulatory element(s) of the *UmPpc* promoter responsible for M cell-specificity and activity are located within the first 500 bp upstream of the translational start codon ATG.

The proximal 500 bp of other C_4 grass *C_4Ppc* promoters are sufficient to drive GUS expression in the M cell of rice

5´-deletion analysis of the *UmPpc* 5´-flanking sequences proved that, the proximal 500 bp are capable to drive M cell-specific expression of the reporter gene in rice leaves. This finding had lead us to test whether the other grass *C_4Ppc* promoters also contain their M cell-specific regulatory elements within their proximal part like *UmPpc*. To determine this, the proximal 500 bp of the *SvPpc*, *PmPpc* and 660 bp of the *ZmPpc* 5´-flanking sequences were fused to the GUS reporter gene and subsequently

transformed in the Nipponbare rice cultivar for analysis (Figure 3A). Due to some unknown complicacy in the cloning procedure of the *ZmPpc* construct, instead of using the proximal 500 bp of the *ZmPpc* promoter we had to use the proximal 660 bp.

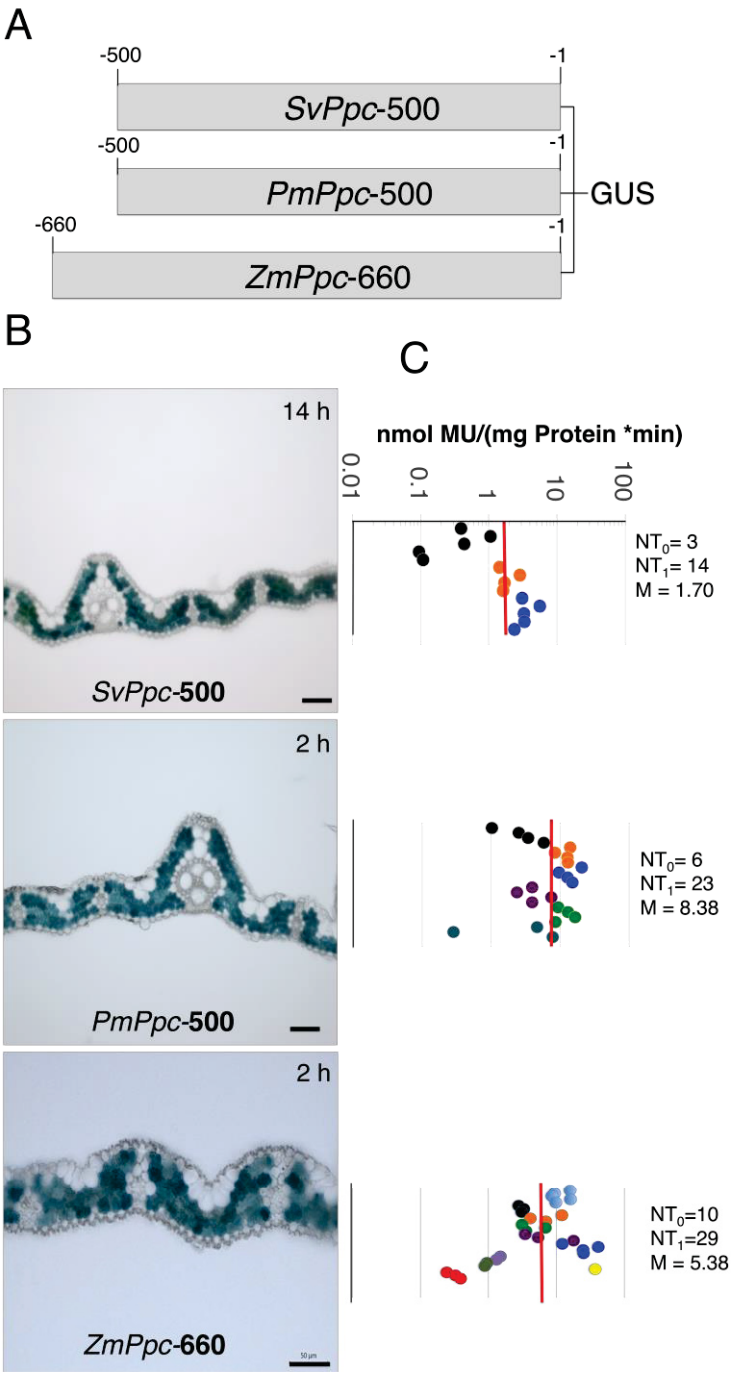


Figure 3: Analysis of the proximal regions of the *C4Ppc* 5´-flanking sequences from *S. viridis*, *P. miliacieum*, and *Z. mays* for promoter activity in rice. (A) Schematic illustration of the chimeric constructs of *SvPpc*-500:GUS, *PmPpc*-500:GUS and *ZmPpc*-660:GUS. **(B)** Histochemical localization of GUS gene expression in the leaf

blades of transgenic rice. Incubation period for *SvPpc*-500:GUS, *PmPpc*-500:GUS and *ZmPpc*-660:GUS was 14, 2 and 2 hours respectively. Scale bar: 50 μm. **(C)** GUS activities in leaves of transgenic rice plants. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number of transgenic plants (NT_1) analyzed from the independent transgenic (NT_0) lines and the median values (M) of GUS activity for each construct are represented at the right side of each column. The same color dots represent the transgenic T_1 plants originate of the same T_0 line. The positions of the median values are marked in red lines.

Histochemical analysis of the *SvPpc*-500:GUS, *PmPpc*-500:GUS and *ZmPpc*-660:GUS constructs revealed that the GUS expression pattern of all three constructs was essentially the same as that of the *UmPpc*-500:GUS construct (Figure 3B). GUS quantitative assays of these three constructs demonstrated that they are also relatedly active as the *UmPpc*-500:GUS construct (Figure 3C). Taking together the data of the *UmPpc*-500, *PmPpc*-500, *SvPpc*-500 and *ZmPpc*-660 constructs provided convincing evidence that the *cis*-regulatory element(s) driving M cell-specificity and activity are located within the proximal regions of the analyzed four grass *C_4Ppc* promoters.

Mapping of the regulatory elements in the proximal regions of the *C_4Ppc* promoters

The data presented so far clearly demonstrated that the proximal 500 bp of our selected C_4 grass *Ppc* promoters contain all the relevant regulatory elements for the M cell-specific expression. In order to gain insight into and to identify the putative *cis* element(s), it would be helpful to have detail map of the grass *C_4Ppc*-500 fragments. To determine whether there are any conserved nucleotide sequences present among the Panicoid C_4 grass *Ppc* promoters, we pursued a comparative analysis.

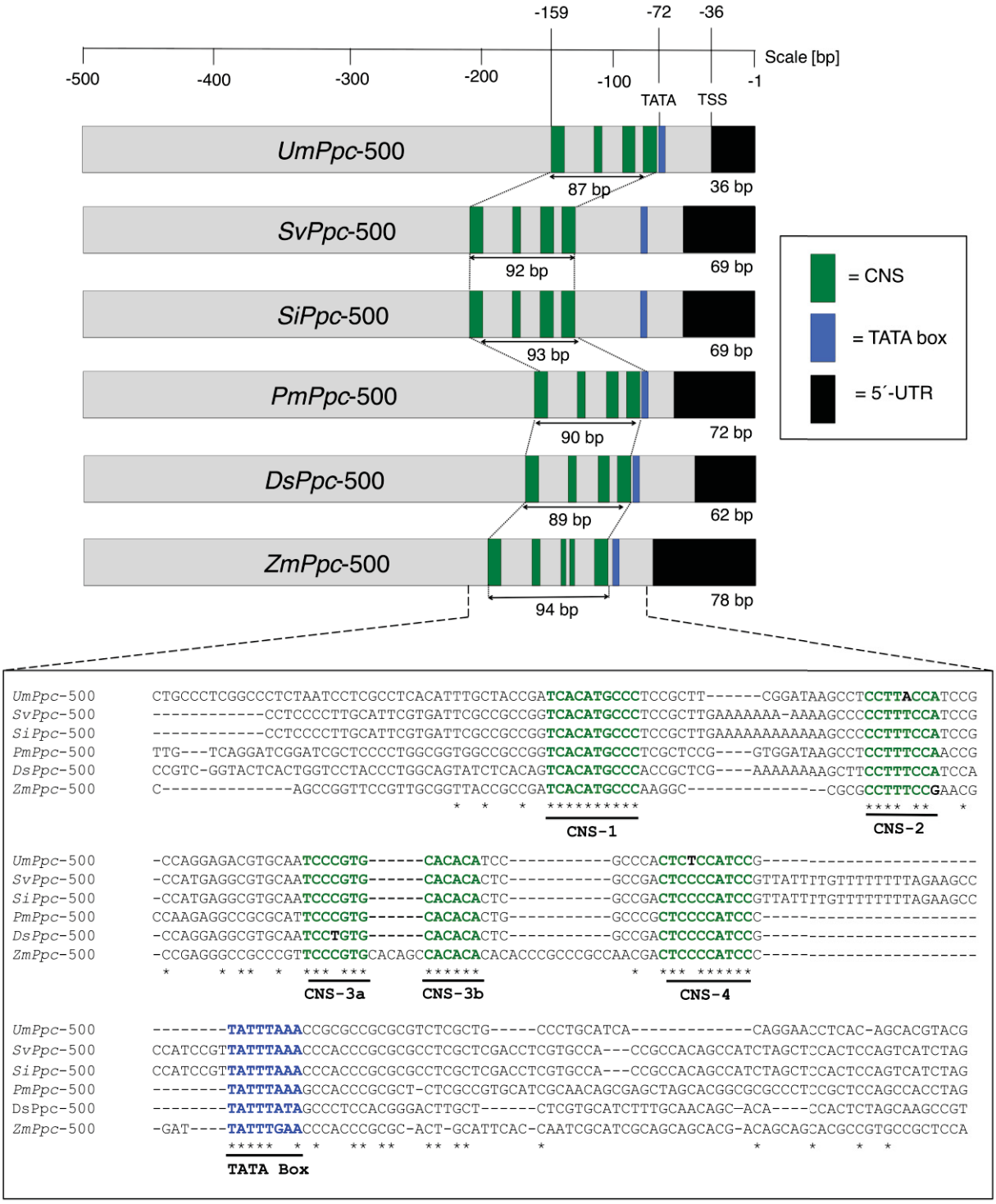


Figure 4: Comparison of the proximal 500 bp of the 5´-flanking regions of the *C4Ppc* genes from the Panicoid C_4 grass species *U. maxima (UmPpc), S. viridis (SvPpc), S. italica (SiPpc), P. miliaceum (PmPpc), Digitaria sanguinalis (DsPpc)* and *Z. mays (ZmPpc)*.

The schematic figure of top of the figure depicts the relative positions of the conserved nucleotide sequences (CNS), putative TATA box and the 5´-untranslated region (5´-UTR). The CNSs are indicated in green, the putative TATA box in blue and the putative

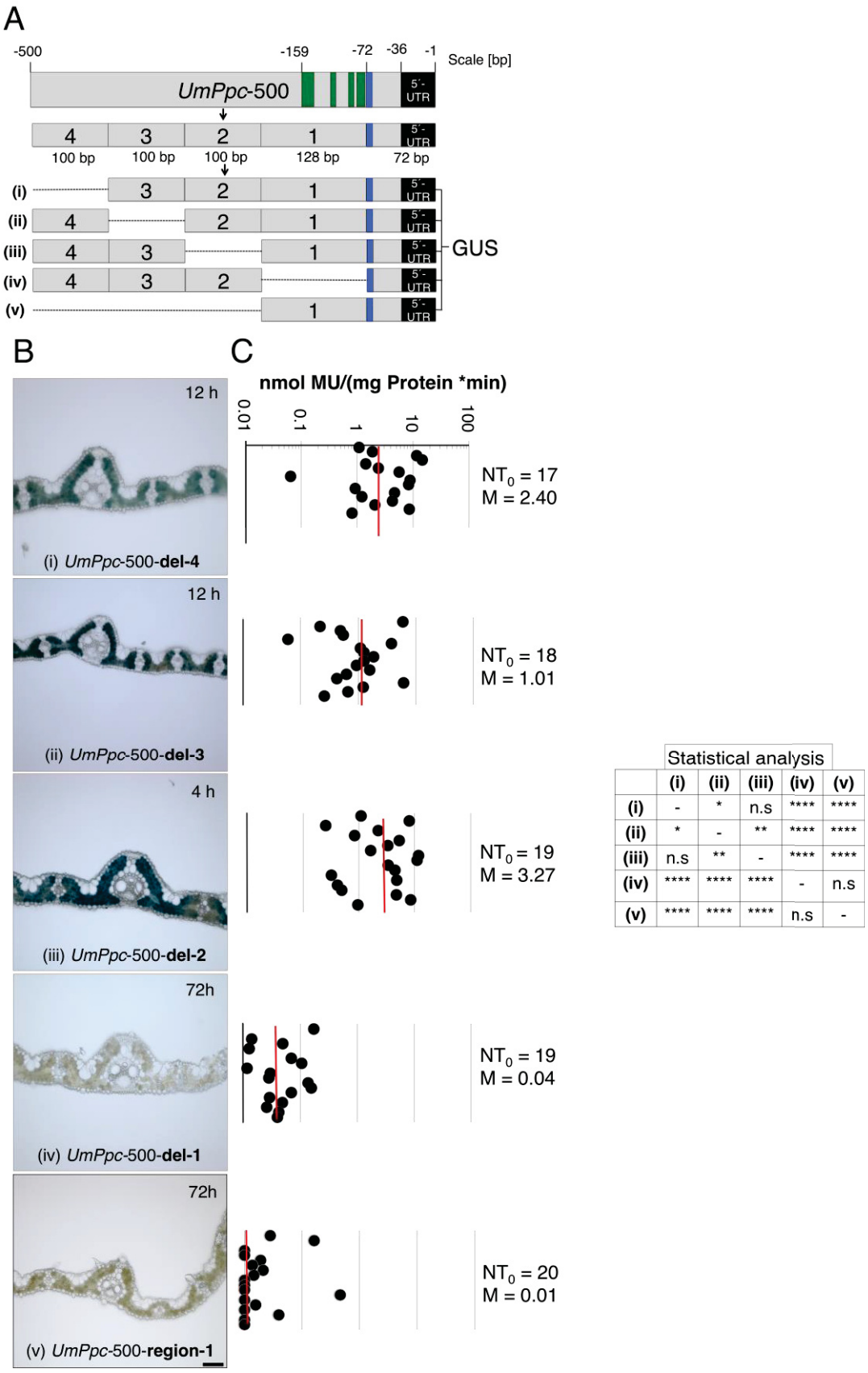
5′-UTR in black boxes. Relative length of the CNS harboring regions and the 5′-UTRs are indicated in bp below of each respective box. Nucleotide numbers refer to the translational start codon ATG (+1) and the positions are related to the black scale bar on top. TSS: putative transcription start site. The asterisk (*) indicates fully conserved base pairs.

In order to carry out the sequence comparison, two additional 5′-flanking sequences of the *C_4Ppc* gene from C_4 panicoid grass *Digitaria sanguinalis* and *S. italica* were also isolated through vectorette PCR (genome walking) (Siebert et al., 1995). Sequence comparison of the proximal 500 bp of the *UmPpc* 5′-flanking sequences with five selected grass *C_4Ppc* 5′-flanking sequences (*Z. mays, S. viridis, S. italica P. miliaceum,* and *Digitaria sanguinalis*) identified four motifs of conserved nucleotide sequences (CNS) (Figure 4). To follow up our experiments we named them as CNS-1, CNS-2, CNS-3 (in case of *ZmPpc,* CNS-3 is divided into CNS3a and CNS3b, due to the insertion of 6 additional nucleotides in the CNS-3) and CNS-4. In the *UmPpc* promoter the four CNSs are located in the range of -159 bp to -73 bp (Figure 4). Downstream to these CNSs, a putative TATA box is also found to be highly conserved in all six *C_4Ppc* 5′-flanking sequences. In the 5′-flanking region of *UmPpc*, the putative TATA box element was detected at the position -72 to -64 bp. Rapid Amplification of 5′-complementary DNA (cDNA) Ends (5′-RACE) analysis was used to define the putative transcription site of the *UmPpc*, *SvPpc*, *SiPpc*, *PmPpc*, and *DsPpc C$_4$Ppc* 5′-flanking sequences. In addition, information about the 5′-UTR regions of the *ZmPpc* was obtained from available Genbank data libraries (Accession No: X15239). Except the CNSs comprising region and the putative TATA box sequence, there is no any other noticeable sequence conservation observed among the six-selected grass *C_4Ppc*-500 fragments (see Supplementary sequence data 5). This observation suggests that the CNSs harboring region could be the putative hot area that plays a crucial role in grass *C_4Ppc* promoters and might carry the M cell-specific determinants. In order to generate a strong argument in favor for our hypothesis it is necessary to produce a deletion construct of the CNSs carrying region.

Deletion of the CNSs comprising sub-fragment from the *UmPpc*-500 fragment causes loss of activity

To unravel the potential role of the identified CNSs, it was indispensable to perform promoter deletion and recombination tests in the analyzed *C_4Ppc* 5′-flanking sequences of the Panicoid grass, and we chose the *UmPpc*-500 fragment as an example. The fragment was sub-divided into four different sub-fragments; 1, 2, 3 and 4, starting upstream of the TATA box (Figure 5A). Sub-fragment-1 contains all the CNSs that were identified in the six selected C_4 grass *Ppc* 5′-flanking sequences. Sub-fragment-2, 3, and 4 do not have any potential conserved nucleotide sequences in the compared six C_4 grass *Ppc* promoters. With the four sub-fragments, five consecutive promoter deletion and recombination constructs were created and designated as *UmPpc*-500-del-4:GUS, *UmPpc*-500-del-3:GUS, *UmPpc*-500-del-2:GUS, *UmPpc*-500-del-1:GUS and *UmPpc*-500-region-1:GUS respectively (Figure 5A). All the five chimeric constructs were stably transformed in Kitaake rice.

GUS histochemical analyses of transgenic rice leaf blade cross-sections showed that the individual removal of the sub-fragment-4, 3, and 2 from *UmPpc*-500 fragment had no effect on spatial accumulation of GUS (Figure 5B). The pattern of GUS expression of these cross sections was essentially same to the GUS expression pattern observed in transgenic rice plants with the *UmPpc*-500 fragment. In contrast, deletion of the sub-fragment-1 from the *UmPpc*-500 fragment resulted complete loss of GUS staining, indicating the sub-fragment-1 is essential for promoter activity (Figure 5B). Surprisingly, the sub-fragment-1 together with its downstream sequences was unable to drive GUS expression in the M cell of rice leaf (Figure 5B). This implies that the presence of additional *cis* regulatory elements upstream of the sub-fragment-1. To correlate the result of GUS histochemical analysis with the activity assay, quantitative measurement of GUS activity was carried out with the leaf extracts of transgenic rice plants. The GUS fluorometric analysis showed almost same type of GUS activity between the transgenic plants of sub-fragment-4 and sub-fragment-2 deletion constructs. However, the transformants of the sub-fragment-3 deletion construct showed 2.4 and 3.2 folds lower GUS activity compared to the GUS activity of the sub-fragment-4 and sub-fragment-2 deletion transformants respectively (Figure 5C), but the



Statistical analysis

	(i)	(ii)	(iii)	(iv)	(v)
(i)	-	*	n.s	****	****
(ii)	*	-	**	****	****
(iii)	n.s	**	-	****	****
(iv)	****	****	****	-	n.s
(v)	****	****	****	n.s	-

Figure 5: Deletion and recombination analysis of the *UmPpc*-500 proximal promoter fragments.

(A) Schematic presentation of the subdivision of the *UmPpc*-500 promoter fragment into four sub-fragments. Sub-fragment-1 (128 bp) consists the all four putative CNSs, while the sub-fragment-2, 3 and 4 (100 bp each) do not have any conserved sequence. Based on the division of sub-fragments, five promoter deletion and recombination constructs: (i) *UmPpc*-500-del-4:GUS, (ii) *UmPpc*-500-del-3:GUS (iii) *UmPpc*-500-del-2:GUS, (iv) *UmPpc*-500-del-1:GUS and (v) *UmPpc*-500-region-1:GUS were prepared. The TATA box and the 5′-UTR are represented in a blue and black box respectively. **(B)** Histochemical localization of GUS (blue staining) in leaf cross sections of transgenic Kitaake rice transformed with the chimeric constructs (i) *UmPpc*-500-del-4:GUS, (ii) *UmPpc*-500-del-3:GUS , (iii) *UmPpc*-500-del-2:GUS, (iv) *UmPpc*-500-del-1: GUS and (v) *UmPpc*-500-region-1:GUS. Incubation times for (i), (ii) and (iii) were 12, 12 and 4 hours respectively and for constructs (iv) and (v) were 72 hours. Scale bar: 50 μm. **(C)** GUS activities in leaves of transgenic rice plants. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number of transgenic plants analyzed (NT_0) and the median (M) values of the GUS activity for each construct are represented at the right side of each column. The positions of median values are marked in red lines. (* $p < 0.05$; ** $p < 0.005$; ** **$p < 0.00005$; n.s., not significant $p > 0.05$).

M cell-specific accumulation of GUS was well maintained in the sub-fragment-3 deletion plants (Figure 5B). This suggests that, sub-fragment-3 might contain a quantitative enhancer element. Conversely, GUS activity in the leaf extracts of sub-fragment-1 deletion transformants is reduced nearly by 60, 30 and 80-fold compared to the GUS activity of sub-fragment-4, 3, and 2 deletion constructs respectively (Figure 5C). A similar GUS activity to sub-fragment-1 deletion construct of the *UmPpc*-500 fragment was also observed with the transformants expressing the *UmPpc*-500-region-1:GUS construct. The quantitative data of *UmPpc*-del-1:GUS and *UmPpc*-500-region-1:GUS constructs completely correlate with the microscopic images of the GUS histochemical analysis. The histochemical and fluorometrical analysis implies that, sub-fragment-1 (128 bp) of the *UmPpc*-500 fragment is essential for driving GUS expression in the M cells of transgenic rice leaves. However, sub-fragment-1 alone is not capable to drive the expression of a reporter gene. It needs further assistance from the upstream sequence to get active. This information leads us to conclude that CNSs

harboring sub-fragment-1 of *UmPpc* is essential for activity but is not sufficient for directing M cell-specific expression.

The CNSs harboring region is also indispensable in the *S. viridis C$_4$Ppc* promoter to drive activity and M cell-specificity

The data of the *UmPpc* promoter proved that the CNSs harboring sub-fragment-1 is absolutely essential for activity and might also be important for M cells-specificity. To test whether this is also the case for other Panicoideae C_4 grass species, a promoter deletion and complementation assay was carried out with the *SvPpc*-500 fragment. First, the putative the CNSs harboring region (92 bp) of the *SvPpc*-500:GUS construct was deleted (Figure 6A). In the second construct the putative CNSs region (92 bp) of *SvPpc* was exchanged with the CNSs region (87 bp) of the *UmPpc* 5´-flanking sequences (Figure 6A). The two chimeric constructs were designated as *SvPpc*-500Δ92:GUS and *SvPpc*Δ92+*Um*87:GUS respectively and subsequently transformed in the Kitaake rice cultivar.

Removal of the 92 bp CNSs region from the *SvPpc*-500:GUS construct reduced the promoter activity almost completely, however, out of 15 transgenic plants four plants produced weak GUS expression in the M cell (Figure 6B) (the respective fluorometric value to the shown cross section is marked with a black arrow in Figure 6C). Nevertheless, the remaining 11 transgenic plants showed extremely low level of GUS activity and did not show blue staining in histochemical assay. We therefore conclude that deletion of the CNSs harboring region from the *SvPpc*-500 fragment reduces the promoter activity at least drastically, if not abolishing activity completely. Complementation of the deleted *SvPpc* CNSs region (92 bp) with the *UmPpc* CNSs containing region (87 bp) regained the activity as well as M cell-specificity of the promoter (Figure 6). Here, we analyzed 15 independent transgenic plants and all the transformants displayed high GUS activity in the M cells (Figure 6B). Quantitative measurement of enzyme activity reveals that the median of the GUS activity of the *SvPpc*Δ92+*Um*87:GUS transformants is almost 100 fold higher than the GUS activity of the *SvPpc*-500Δ92:GUS construct (Figure 6C). We conclude from this deletion and heterologous substitution experiments that the putative CNS are obligatory for *C$_4$Ppc* promoter activities of the analyzed C_4 Panicoid grass species. The conservation of the

CNSs in other C_4 grasses of the Panicoids suggests their general importance in *C_4Ppc* promoter function.

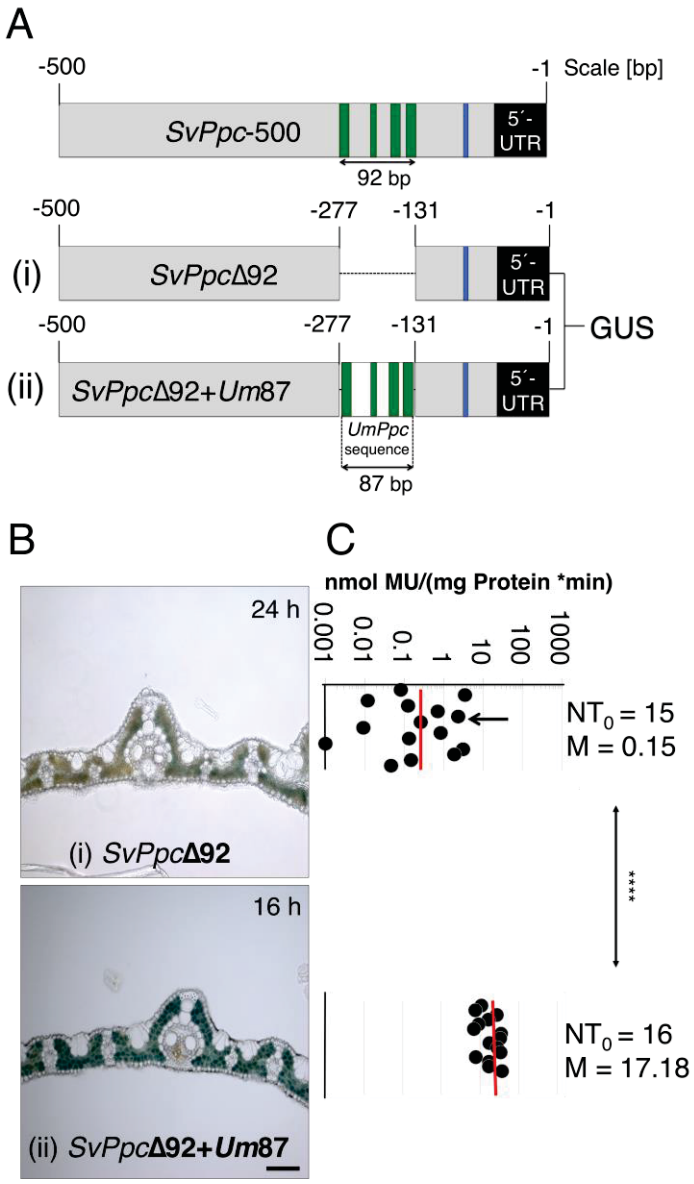


Figure 6: Deletion and complementation analysis of CNSs containing region of the *SvPpc*-500.

(A) Overview of the chimeric deletion and complementation constructs of the CNSs harboring region of the *SvPpc*-500:GUS construct. **(B)** Histochemical analysis of transgenic rice plants transformed with *SvPpc*-500Δ92:GUS and *SvPpc*Δ92+*Um*87:GUS respectively. Scale bar: 50 µm. Incubation times were 24 hours and 18 hours for *SvPpc*-500Δ92:GUS and *SvPpc*Δ92+*Um*87:GUS transformants respectively. **(C)** GUS activities are expressed in nanomoles of the reaction product 4-

methylumbelliferone (MU) per mg of protein per minute. The number (NT_0) of independent transgenic plants analyzed and the median (M) values of GUS activity for each construct are represented at the right side of each column. The positions of median values are marked in red lines. (**** $p < 0.00005$).

The nucleotide sequence between the CNSs does not have any impact on directing activity and M cell-specificity

The comparative analysis of the six-selected *C_4Ppc* 5´-flanking sequences showed there is some degree of conservation of the nucleotide sequences between the sequences of the four identified CNSs (Figure 4). One might speculate that, not only the four putative CNSs are important, but also the adjoining nucleotides may have an impact on directing the activity and specificity. To address this, the nucleotide sequences between CNS-1 and CNS-2 (18 bp), and CNS-2 and CNS-3 (19 bp) of the *UmPpc*-500-del-4:GUS constructs were replaced with some random foreign nucleotide sequences (nucleotide sequence from the GFP coding gene). The resulting construct *UmPpc*-fGFP:GUS was examined in Kitaake rice (Figure 7).

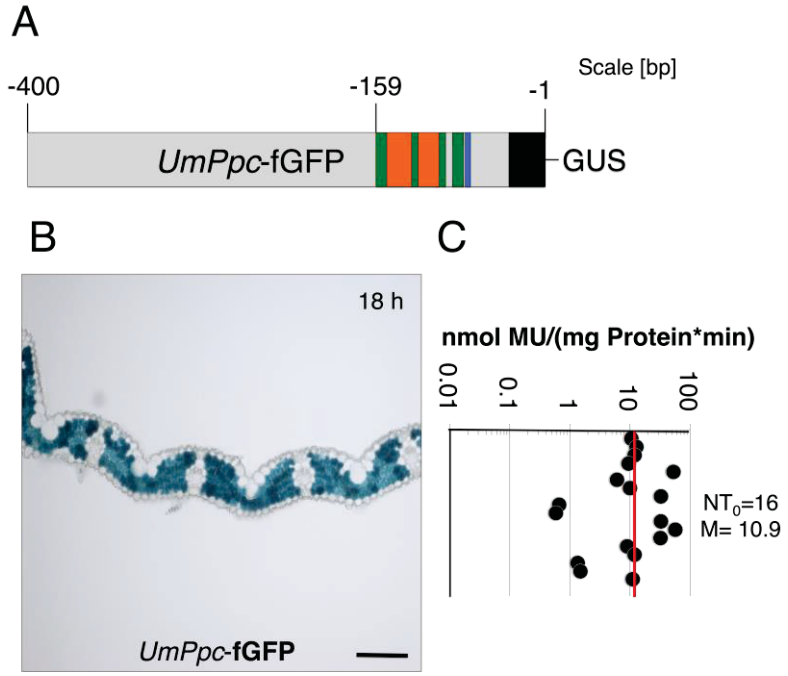


Figure 7: Functional analysis of the *UmPpc*-fGFP promoter in leaves of transgenic rice.

(A) Schematic structure of the *UmPpc*-fGFP:GUS construct. The orange boxes represent the foreign nucleotide sequences (sequence from the GFP coding gene). **(B)**

Histochemical localization of GUS in a leaf cross section of transgenic rice transformed with the *UmPpc*-fGFP:GUS construct. Incubation period in GUS staining buffer was 12 hours. Scale bar: 50 µm. **(C)** GUS activities in leaves of transgenic rice plants. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number (NT_0) of transgenic plants analyzed and the median (M) value of the GUS activity for the construct are represented at the right side of the column. The position of median value is marked as a red line.

GUS histochemical and quantitative analyses of 16 independent transgenic plants of the *UmPpc*-fGFP:GUS construct showed a comparable GUS expression to the *UmPpc*-500-del-4:GUS construct (Figure 5). This indicates that the nucleotide sequences between the CNSs sequence has no major impact in directing activity and M cell-specificity.

The TATA and 5′-UTR segment of the *UmPpc* promoter can be replaced with a minimal 35S promoter

It was tempting to speculate, whether the TATA box and 5′-UTR segment (72 bp) of the *UmPpc* 5′-flanking sequences contains any important *cis* regulatory element(s) and if it has any effect on directing M cell-specific expression. Therefore, we replaced this 72 bp sequence from the *UmPpc*-500-del-4:GUS construct (Figure 5) with the -60 bp minimal 35S promoter (M35S) of the cauliflower mosaic virus, resulting in the formation of *UmPpc*-M35S:GUS construct (Figure 8A). In *Arabidopsis thaliana* the minimal 35S promoter alone cannot drive expression of GUS gene (data not shown) and therefore we used this -60 bp 35S promoter fragment to replace the TATA box and 5′-UTR segment (72 bp) of the *UmPpc* 5′-proximal region.

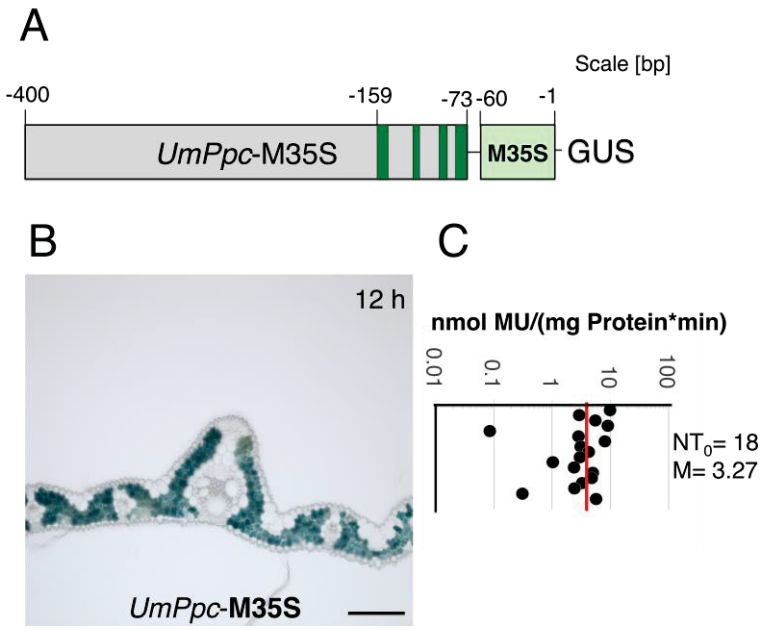


Figure 8: Functional analysis of the TATA and 5´-UTR segment of the *UmPpc* promoter.

(A) Schematic presentation of the *UmPpc*-M35S:GUS construct. TATA box and 5´-UTR segment of the *UmPpc*-500-del-4:GUS construct was replaced with the -60 minimal 35S promoter (M35S) sequence. Here, the M35S sequence is indicated in light green box. **(B)** shows the GUS histochemical results of cross section of a rice leaf blade from a plant transformed with the *UmPpc*-M35S:GUS construct. Incubation period in GUS staining buffer was 12 hours. Scale bar: 50 µm. **(C)** represents the MUG activity assay data of *UmPpc*-M35S:GUS leaf extract. GUS activities are expressed in nanomoles of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number (NT_0) of transgenic plants analyzed and the median (M) values of GUS activity are represented at the top of column. The position of median values is marked in red line.

After examining 18 independent transgenic rice plants of the *UmPpc*-M35S:GUS construct, we did not find any substantial difference in terms of activity and M cell-specificity between the *UmPpc*-M35S: GUS construct (Figure 8) and the *UmPpc*-500-del-4:GUS (Figure 5). This gives a clear indication that neither the putative TATA box nor the 5´-UTR of the *UmPpc* promoter contain any relevant *cis* element(s) for directing activity and M cell-specificity.

(iv) Discussion

C_4 photosynthesis is a complex biochemical trait, which has evolved over 60 times in the angiosperm and at least 22 times independently in the grasses (Sage, 2004; Sage, 2016). A cross-species expression specificity experiment performed by Matsuoka and his colleagues (1993) showed that the *C4Ppc* promoter of maize maintained its M cell-specific expression in the transgenic C_3 rice. This implies that the gene regulatory system of rice and maize share common *trans*-regulatory factors that recognize cell-specific *cis*-regulatory modules in the C_4 maize promoter and interpret them correctly in rice. A complete molecular phylogenetic tree developed for the grass family showed that all the C_4 grass species are clustered in the PACMED clade (Christin et al., 2008; Christin et al., 2009). Panicoideae is a subfamily of PACMED clade, which contains many of the commercially important C_4 grasses such as maize (*Zea mays*), pearl millet (*Pennisetum glaucum*), sorghum (*Sorghum bicolor*), sugar cane (*Saccharum officinarum*), and foxtail millet (*Setaria italica*) (Kellogg, 2001). In contrast to the C_4 grasses, the phylogenetic location of the C_3 rice is in the BEP clade. In an evolutionary context of view, the two group of grasses of PACMED-BEP have diverged more than 50 million years before (Christin et al., 2008). Despite being separated for a longer period of time, it is quite surprising that maize and rice share common *trans*-regulatory element(s) for M specific cell-specific expression of *C4Ppc* gene. One could assume that the common *trans*-regulatory element(s) that are necessary for M cell-specific expression of *C4Ppc* gene already existed in a common ancestor of PACMED and BEP clades. In order to generate a strong argument in favor of this hypothesis, we isolated the *C4Ppc* 5′-flanking sequences from three further C_4 Panicoid grass species (*S. viridis*, *P. miliaceum*, and *U. maxima*) of PACMED clade and subsequently analyzed their promoter potentiality in rice. Surprisingly, all the three *C4Ppc* 5′-flanking sequences were found to maintain their C_4 type cell-specific expression when tested in rice (Figure 1). This finding supports the proposed hypothesis of Matsuoka et al., (1994). Like the grasses, a similar type of evolutionary conservation of the transcription regulatory system of BS cells can also be observed in distantly related dicotyledonous C_3 and C_4 plants. A previous cross-species expression study with the C_4 *Flaveria trinervia GLDPA* promoter in the C_3 *Arabidopsis thaliana* showed that the promoter is active and BS cell-specific in Arabidopsis (Engelmann et al. 2008; Wiludda et al. 2012).

Conversely, the promoter of the sulphate transporter gene (*SULTR2;2*) from Arabidopsis has been found to maintain its BS cell-specificity in the Asteracean C_4 species *Flaveria bidentis* (Kirschner et al., 2018). These findings indicated that *trans*-regulatory elements of BS cells are at least partially evolutionary conserved in the dicotyledons plants. Therefore, the evolutionary conservation of the *trans*-regulatory systems of M cells in the C_3 and C_4 grass species is not an uncommon phenomenon.

One of the major interests of our present study was to define the minimal promoter of *C_4Ppc* that is capable of directing similar patterns of M cell-specific expression as their full-length 5´-flanking sequences. Previously, it was reported from the chromatin organization point of view that the promoter of *ZmPpc* is composed of a proximal part (up to -600 bp of ATG start codon) and a distal part (upstream of -2 Kb) (Offermann et al., 2006; Offermann et al., 2008). M cell-specific expression of *ZmPpc* was assumed to be associated with the methylation of lysine-4 of histone-3 (H3K4) in the proximal region of *ZmPpc* promoter nucleosomes (Danker et al., 2008). In contrast, acetylation in the distal part of promoter nucleosomes was considered to play an important quantitative role for transcriptional regulation (Offermann et al., 2008). In our study, analysis of the 5´-deletion experiments of the *C_4Ppc* 5´-flanking sequences of *ZmPpc*, *SvPpc*, *PmPpc*, and *UmPpc* showed that about 500 bp of 5´-flanking sequences (with respect to the ATG start codon) are sufficient for full mesophyll promoter activity, i.e. all the *cis*-regulatory elements required for promoter strength and specificity are contained in about 450 bp (upstream to the 5´-UTR). This finding also indicates that the M cell-specific regulatory mechanism of *C_4Ppc* gene of Panicoid grass is different from the *ppc*A gene of the C_4 Asteracean species *Flaveria trinervia*, albeit both the genes were experimentally proven to be regulated at a transcriptional level (Matsuoka et al., 1994; Gowik et al., 2004). The M cell-specific expression of the C_4 *ppc*A promoter of *F. trinervia* in the transformable, closely related C_4 species *F. bidentis* was found to be regulated by the two important regions, a proximal fragment up to -570 bp (with respect to the ATG start codon) and a distal region from -1551 to -2141 bp. The proximal part of promoter is mainly responsible for the quantitative expression of the *ppc*A gene. The distal part of the promoter contains a 41 bp enhancer like expression module-1 (MEM-1) and thought to be responsible for the M cell-specific expression of C_4 *ppc*A gene by suppressing its expression in the BS cells. In contrast to the C_4 *ppc*A promoters of dicot *Flaveria*, our analyzed *C_4Ppc* promoters of Panicoid grasses contain

their all regulatory elements in the proximal part of their promoters. This data clearly suggests that the length of the *ppc*A promoter of C_4 *Flaveria* is larger than our analyzed *C_4Ppc* promoters of the Panicoid grasses. A cap analysis of gene expression (CAGE) experiment performed by Mejía-Guerra et al., (2015) predicted that many of the *Z. mays* genes, but by no means all, have sharp transcription start site clusters (TC). In addition to that, a total of 38 % promoters with the sharp TC have been reported to possess a broader TATA consensus motif and are assumed to associate with the tissue-specific expression of those respective genes (Mejía-Guerra et al., 2015). A genome-wide chromatin immunoprecipitation experiment performed in the dicot plant *Arabidopsis thaliana* showed that the TFs bind with the highest frequency close to the transcription start site (TSS) of their target gene promoters (Heyndrickx et al., 2014). Concerning the position of TFs binding sites with respect to the gene, it is assumed that the core promoter elements of most maize gene are located up to 1 Kbp upstream of their TSS (Mejía-Guerra et al., 2014). A further study of a gene regulatory network for the phenolic pathway has provided enough evidence to support this assumption. It was reported that the transcription factors (TFs) regulating the expression of 54 key genes of maize phenolic pathway bind in the promoter region up to 1 Kb upstream of the TSS (Yang et al., 2017). Taken together these results indicate that the length of most promoters of maize genes could be relatively small and our findings with the *C_4Ppc* promoters completely supports this hypothesis. However, it is still remained unclear, which selective pressure is responsible for making the promoter sizes of maize and other Panicoid grass species genes relatively short.

To localize the regulatory elements of the panicoid grass *C_4Ppc* promoter, we compared the proximal 500 bp of *C_4Ppc* 5′-flanking sequences from the six different Panicoid C_4 grass species, including the *Z. mays*, for which genome sequences are available. From the sequence comparison analysis, we observed that *C_4Ppc* 5′-flanking sequences of the six selected Panicoid grasses share four motifs of CNSs. To the best of our knowledge, the role of these CNSs on the grass *C_4Ppc* promoters is still elusive. Downstream to these CNSs, a putative TATA box can also be detected in all the six analyzed 5′-flanking sequences. In metazoan and *Drosophila melanogaster*, it has been reported that the promoters (Type-I promoters) possessing a TATA motif are usually involved in tissue-specific expression of the respective genes (Engstrom et al., 2007; Lenhard et al., 2012). The presence of a putative TATA motif in our analyzed

promoters is, therefore, supports the criteria of being a cell-specific promoter. Since the proximal part of the *ZmPpc*, *SvPpc*, *PmPpc*, and *UmPpc* promoters gave comparable GUS expression in transgenic rice leaf, the existence of conserved TATA box and CNSs in their sequences indicate there might be a common regulatory mechanism among these *C4Ppc*. The most parsimonious explanation is that all the four proximal part of *C4Ppc* 5′-flanking sequences share a common characteristic, which is recognized by the same *trans* factor(s) of rice to generate M cell-specificity.

The necessity of the CNSs for activity was proven by the deletion of the region containing CNSs from the *UmPpc-500* and *SvPpc*-500 (Figures 5 and 6). This gives an indication about the existence of a quantitative module in the CNSs harboring region. Interestingly, all the four CNSs were also detected in the proximal 500 bp of a *ppc* homolog gene (*ppc1*) from the C_3 Panicoid grass *Dichanthelium oligosanthes* (Supplementary Figure 1). However, in comparison to the expression of *C4Ppc* of C_4 Panicoid grass, the expression level of the *ppc1* gene in *D. oligosanthes* was almost undetectable (Studer et al., 2016). This indicates that the CNSs are not the only players for regulating the M cell-specific expression of *C4Ppc* gene of Panicoid grass. This assumption was justified when we identified that the CNSs harboring region—together with its basal sequences (TATA and 5′-UTR segment) of the *UmPpc* promoter—was not sufficient to drive GUS expression in the M cells of a rice leaf (Figure 5). This implies the presence of additional regulatory elements in the upstream of the CNSs, which assist the CNSs at driving activity and M cell-specificity. This was verified by using the *SvPpc-500*Δ92 and *SvPpc*Δ92+*Um*87 constructs (Figure 6). Upon deletion of the CNSs harboring region of the *SvPpc*-500 fragment, 75 % of the total transgenic rice plants lost the complete activity of GUS in the M cells, while only 25 % plants showed a weak GUS expression in the M cells. This implies, in addition to the CNSs region, some other assisting regulatory element(s) exist within the 500 bp fragment, which are capable of partially complementing the deletion effect of the CNSs. Further, we investigated whether a plausible interaction between the CNSs and the assisting element(s) is necessary. This hypothesis was justified by using *UmPpc*-500-del-1, *UmPpc*-500-region-1, *SvPpc*-500Δ92, and *SvPpc*Δ92+*Um*87 constructs (Figures 5 and 6). In the *UmPpc*-500-del-1 and *UmPpc*-500-region-1 constructs, the assisting element(s) and the CNSs were separated in the two constructs respectively; both the constructs failed to produce GUS expression in the rice leaf. This implies the

importance of the interaction between the CNSs and unknown assisting element(s). Failure of the interaction between the CNSs and the assisting elements caused nearly or complete loss of activity (Figures 5 and 6). We assume that the assisting element(s) to the CNSs region are redundant in nature. A consecutive deletion of the sub-fragment-4, 3 and 2 independently from the *UmPpc*-500 fragment did not abolish the GUS gene expression in the M cell (Figure 5). If the assisting element(s) were not redundant, then it was expected to be located in any of the three marked sub-fragments and deletion of that respective sub-fragment would produce a substantial difference in the activity and specificity. However, the elimination of the sub-fragment-3 showed a slight drop in activity in compared to the activity of sub-fragment-4 and sub-fragment-2 deletion constructs (Figure 5). It could be possible that the sub-fragment-3 consists of an enhancer element. Since the activity and specificity was not lost completely when the sub-fragment-3 was deleted and therefore, we couldn't conclude that this is the only sub-fragment that contains the assisting elements to the CNSs. Moreover, no specific repeated element(s) in the nucleotide sequences were detected, which are shared across the sub-fragment-2,3, and 4. It has already been proven in many studies that a cell-specific regulator can also be redundant in nature without possessing identical or repeated sequences (Hatton et al., 1995; Kirschner et al., 2018). The phenylalanine ammonia-lyase-2 (*PAL2*) promoter of *Phaseolus vulgaris* L. and the sulphate transporter-2 (*AtSULTR*2;2) promoter of Arabidopsis are possibly the best examples of such a scenario (Hatton et al., 1995; Leyva et al., 1992; Kirschner et al., 2018). The (*PAL2)* promoter confers xylem-specific expression in tobacco plant. A detailed investigation revealed that the nucleotide sequences range from -289 to -74 bp of the *PAL2* promoter is essential for the xylem-specific expression of a reporter gene. The deletion of the proximal or distal part of that region did not affect the expression pattern of the reporter gene, confirming the redundancy of the xylem-specific elements (Leyva et al., 1992; Hatton et al., 1995). In the *AtSULTR2;2* promoter, a sequence consisting of 350 bp (sequence ranges from -1845 to -1495 bp) was found to drive BS cell and vein specific expression of a reporter in Arabidopsis (Kirschner et al., 2018). Further consecutive deletions analysis within this 350 bp segment failed to eliminate the spatial patterning of the reporter gene, suggesting the redundant nature of the regulatory elements (Kirschner et al., 2018). Hence, the redundancy of a regulatory element in the promoter is not an unusual incidence.

(v) Conclusion

The complexity of the C_4 photosynthetic metabolism represents enormous challenges for installing the C_4 pathway into C_3 rice plant. The discovery of a new set of M cell-specific promoters is, therefore, a key finding for the C_4 rice engineering. In addition to that we found that around 450 bp (upstream of their TSS) of our analyzed promoters contain all the essential regulatory elements for driving M cell-specific expression in rice leaf. Since the C_4 rice engineering might require transformation of array of genes, therefore, using the shorter size of promoter would be helpful to keep the constructs size in a considerable and thus minimize the complicacy of rice transformation method. Importantly, four motifs of CNSs were identified in our analyzed promoters, which are essential for the activity of the promoter. Identification of the CNSs and its interaction with the unidentified redundant upstream assisting element(s) resolve the unknown mystery of the *C_4Ppc* regulatory mechanism of the multiple closely related C_4 grass species. Thereby, the identified CNSs could be defined as the putative regulatory element(s) for directing activity of the grass *C_4Ppc* genes. Moreover, the redundant nature of the upstream assisting element(s) provides a scope to reduce the length of the analyzed promoters even shorter in near future. We hope that the data presented herein will encourage others to develop synthetic M cell-specific promoters that will be useful for developing C_4 rice.

(vi) Materials and Methods

Rapid Amplification of 5′-cDNA Ends (5′ RACE-PCR):

The 5′-UTR of the *SvPpc*, *PmPpc*, *UmPpc* and *DsPpc* were determined using the 5′-RACE PCR following to the manufacturer′s manual of either the SMART cDNA synthesis Kit or the SMARTer™ RACE cDNA Amplification Kit (Clontech Laboratories, Mountain View, USA). At the first stage of this process, total RNA from the leaves of respective wild type plants of *S. viridis*, *P. miliaceum*, *U. maxima* and *D. sanguinalis* were extracted with the RNeasy® Plant Mini Kit (Qiagen, Hilden, Germany). 1µg RNA was used for the cDNA first strand synthesis in all cases. Afterwards, the PCR amplification of the 5′-UTRs was performed using the

Advantage® 2 DNA Polymerase Mix (Clontech Laboratories, Mountain View, USA) or the Phusion HF DNA Polymerase (Thermo Fisher, Waltham, USA). The 3´-gene specific oligonucleotides that were employed in combination with the either the SMART-II-A 5´-Primer or the SMARTer^{TM} Universal Primer A Mix (UPM) in the 5´-RACE PCR are listed in the Table-1.

Table 1: List of oligonucleotides used to determine the 5´-UTR sequences

Species	RACE 5´-oligonucleotides	RACE 3´ gene specific oligonucleotides (5´to 3´)	Kit used
Setaria viridis	UPM (Universal Primer A Mix)	*viridis*RACE1: TCAGCAGTGGCCGCTTGCCG	SMARTer^{TM} RACE cDNA Amplification Kit
Digitaria sanguinalis	UPM	*sanguinalis*RACE1: CAGCTGGAGTGCACGTTCGCGTGTGT	SMARTer^{TM} RACE cDNA Amplification Kit
Urochloa maxima	SMART-II-A 5´-Primer	*maxima*RACE1: AGTCGGAGTAGCCGACCATC	SMART cDNA Synthesis Kits
Panicum miliaceum	SMART-II-A 5´-Primer	*miliaceum*RACE1: GTAGGGGCCGAAGCTGTCGCGG	SMART cDNA Synthesis Kits

The resulting products of the 5´-RACE PCR were cloned into the pJET1.2/Blunt vector (Thermo Fisher, Waltham, USA) and sequenced (LGC Genomics, Berlin, Germany). The 5´-UTR sequences of *ZmPpc* (Accession No: X15239) and *SiPpc* (Accession No: AF495586) were taken from the information present in GenBank/EMBL data libraries.

Isolation of *C$_4$Ppc* 5´-flanking sequences of *S.viridis*, *S.italica*, *P. miliaceum*, *U. maxima* and *D. sanguinalis*:

The 5´-flanking regions of *C$_4$Ppc* genes of *S. viridis*, *S. italica*, *P. miliaceum*, *U. maxima* and *D. sanguinalis* were isolated from total DNA by vectorette PCR (Siebert et al., 1995) using the Universal Genome Walker Kit (Clontech Laboratories, Mountain View, USA) and following the protocol of manufacturer's. For each plant species DNA libraries with the restriction enzymes *Dra*I, *Eco*RV, *Pvu*II, *Stu*I, *Ehe*I, *Ksp*AI, *Sma*I *Smi*I and *Afe*I (Thermo Fisher Scientific, Waltham, USA) were prepared. For the primary and secondary PCR of the 1st walking step, the gene specific oligonucleotides

(GW1 and GW2) were designed that bind specifically to the 5′- end of the coding sequence of *C4Ppc* gene of the respective species (Supplementary Table 1). Together with the adapter primers AP1and AP2 the gene specific primers were used for the 1st walking step. The sequence information obtained in this step was then used for a subsequent 2nd walking step and the next gene specific oligonucleotides (GW3 and GW4) were designed accordingly (Supplementary Table 1). Subsequent walking steps were continued till the desirable length of the 5′ -flanking sequences was achieved for each of the *C4Ppc* gene.

The resulting PCR products from the Genome walking were cloned into pJET1.2/Blunt vector using the CloneJET PCR cloning Kit (Thermo Fisher, Waltham, USA) following the protocol of manufacturer's. Afterwards the cloned plasmids were transferred into chemically competent *Escherichia coli* DH5α cells (Invitrogen™, Carlsbad, USA) and the positive clones were identified via colony PCR using the pJET1.2 Forward Sequencing Primer and the pJET1.2 Reverse Sequencing Primer of the Kit. Next, the plasmids from the positive colonies were extracted using the QIAprep® Spin Miniprep Kit (Qiagen, Hilden, Germany). Finally, the 5′-flanking sequences from the respective plasmids were determined by means of Sanger sequencing. In order to identify the CNSs among the 5′-flanking sequences, multiple sequence alignment was carried out using the online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Cloning of promoter-reporter gene constructs

All the chimeric promoter-reporter gene constructs used in this study were cloned into a slightly modified Gateway pMDC164 vector (Curtis & Grossniklaus, 2003). In the original pMDC164 vector, the marker gene hygromycin phosphotransferase was expressed under control of the cauliflower mosaic virus 35S promoter. It was reported that the strong 35S promoter commonly used in a selectable marker gene of plant transformation vector could affect the expression pattern of the adjacent tissue specific gene-promoters (Zheng et al., 2007). Thereby, in our pMDC164-modified vector, the 35S promoter of the hygromycin cassette has been replaced with a ubiquitin promoter of *Z. mays*. In this study, the full-length promoter and the deletion constructs were generated using the Gateway cloning technique (Hartley et al., 2000). The gateway *att*B

sites were added with PCR to the respective sequences. Oligonucleotide combinations that have been used for PCR amplification of the promoters' sequence are listed in the Supplementary Table 2 and 3. Next the purified PCR products were cloned into the entry vector pDONR207 or pDONR221 following the instruction of the Gateway® BP-Clonase II enzyme mix manual (Invitrogen, Thermo Fisher Scientific, USA). The resulting entry clones were verified by sequencing and then subjected for LR reaction to the destination vector pMDC164-modified. LR reaction was performed according to the manual instruction of the Gateway® LR-Clonase II enzyme mix (Thermo Scientific, Waltham, USA). Positive clones from the LR reaction were verified by means of restriction enzyme test digestion. Internal promoter deletion and recombination constructs (listed in supplementary table 3) were synthesized by Biomatik using the Custom Gene Synthesis service (Biomatik, Ontario, Canada) (http://www.biomatik.com/services/gene-synthesis-9.html). Gateway overhang *att*L sides were added during the synthesis of the constructs and the company cloned them in pUC57 as an entry vector. Next, the synthesized constructs from the pUC57 entry clones were delivered to pMDC164-modified destination vector via LR reaction.

Transformation of *Oryza sativa*

Two different cultivars of *Oryza sativa ssp. japonica* rice, Kitaake and Nipponbare, were used for the transformation in this project. The chimeric promoter-reporter gene constructs that were transformed in the respective Kitaake and Nipponbare rice cultivars are listed in the (Supplementary Table 4). The constructs of interest for the transformation in the Kitaake and Nipponabare rice, were first transformed into the *Agrobacterium tumefaciens* strain AGL1 (Lazo et al., 1991) and EHA105 (Hood et al., 1993) respectively via electroporation. Next callus induction, transformation, selection and, regeneration and rooting of rice were done according to the slightly modified protocol of (Toki et al., 2006). The adopted protocol can be found at https://langdalelab.files.wordpress.com/2015/07/kitaake_transformation_2015.pdf. Hygromycin resistant rice plantlets with well-developed roots from the *in vitro* culture were transferred to hydroponic culture in the green house for two weeks according to protocol described in (Gregorio et al., 1997). Afterwards, the transgenic rice plants were grown to maturity in soil (J. Arthur Bower's, JOHN INNES No.2, Sinclair Pro,

Cheshire, UK) at 30ºC. Finally, the presence of the T-DNA in the transgenic T_0 and T_1 plants was confirmed by genotyping via polymerase chain reaction (PCR)

GUS histochemical and MUG quantitative assay

Mature leaves from positively PCR tested transgenic plants were harvested for histochemical analysis and quantitative GUS fluorometric assay. Regenerated T_0 plants of the Kitaake rice cultivar and T_1 plants of the Nipponbare cultivar harboring the respective constructs were used for the analyses. Fully developed 2nd or 3rd leaves from the top of plant with a length of at least 10 cm were harvested before the onset of flowering. For the histochemical analysis, thin cross sections of mature leaves of the transgenic plants were prepared using a razor blade. Then the cross sections were incubated in a reaction tube with 100 - 200 µl incubation buffer (100 mM Na-Phosphate pH 7.0, 10 mM EDTA, 0.5 mM $K_4Fe(CN)_6$, 0.5 mM $K_3Fe(CN)_6$, 0.1 % (v/v) Triton X-100, 1 mM X-Gluc (5-bromo-4-chlor-3-Indoyl-β-D Glucuronid)) at 37°C for 2 to 72 hours (Engelmann et al., 2008). After removing the incubation buffer, the reaction was stopped by adding 100 µl fixation buffer (75 % (v/v) Acetic acid, 25 % (v/v) Ethanol). The cross sections were rinsed in 70 % ethanol to remove the chlorophyll. Finally, the cross sections were analyzed, and images taken with a Carl Zeiss Axiopot equipped with Axio Cam ICc5. Fluorometric quantification of the GUS activity (MUG assay) was performed according to the protocol of (Jefferson et al., 1987) and (Kosugi et al., 1990). Activity of GUS expression was measured with the help of a plate reader (Synergy/HTX-multi mode reader, BioTek, Vermont, USA). To determine the statistical difference between data sets a Mann-Whitney U test (Prism 6, Graph Pad Software, La Jolla, USA) was employed.

(vii) Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (DFG/GRK1525) through the International Graduate School for Plant Science (iGRAD-Plant) "The dynamic response of plants to a changing environment".

(viii) Accession number

Sequence data presented in this article has been submitted to GenBank data library and are in processing for publication under following accession numbers: MH675618

(*SvPpc* 5′-flanking sequences), MH675617 (*PmPpc* 5′-flanking sequences), MH675616 (*UmPpc* 5′-flanking sequences), and MH675615 (*DsPpc* 5′-flanking sequences)

(ix) Supplementary data

Supplementary Figure: Supplementary Figure 1

Supplementary Tables: Supplementary Tables 1, 2, 3, and 4

Supplementary sequence data: Supplementary sequence data 1, 2, 3, 4, and 5

(x) References

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., & Westhoff, P. (2007). Evolution and Function of a cis-Regulatory Module for Mesophyll-Specific Gene Expression in the C4 Dicot *Flaveria trinervia*. *The Plant Cell Online*, ***19***(11), 3391–3402.

Assaad, F. F., Tucker, K. L., & Signer, E. R. (1993). Epigenetic repeat-induced gene silencing (RIGS) in Arabidopsis. *Plant Molecular Biology*, ***22***(6), 1067–1085.

Brautigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., … Weber, A. P. M. (2011). An mRNA Blueprint for C4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C3 and C4 Species. *Plant Physiology*, ***155***(1), 142–156.

Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., … Langdale, J. A. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C_4 photosynthesis. *Science (New York, N.Y.)*, ***331***(6023), 1436–1439.

Chan, V., Dreolini, L. F., Flintoff, K. A., Llyod, S. J., & Mattenley, A. A. (2002). The effect of increasing plasmid size on transformation efficiency in Escherichia coli. *Journal of Experimental Microbiology and Immunology*, ***2***(April), 207–223.

Chollet, R., Vidal, J., & O'Leary, M. H. (1996). Phosphoenolpyruvate carboxylase: A Ubiquitous, Highly Regulated Enzyme in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, ***47***(1), 273–298.

Christin, P.A., Salamin, N., Kellogg, E. A., Vicentini, A., & Besnard, G. (2009).

Integrating Phylogeny into Studies of C4 Variation in the Grasses. *Plant Physiology*, ***149***(1), 82–87

Christin, P. A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V., & Salamin, N. (2008). Oligocene CO_2 decline promoted C4 photosynthesis in grasses. *Current Biology*, ***18***(1), 37-43.

Covshoff, S., & Hibberd, J. M. (2012). Integrating C 4 photosynthesis into C 3 crops to increase yield potential. *Current Opinion in Biotechnology*, ***23***(2), 209–214.

Curtis, M. D., & Grossniklaus, U. (2003). A Gateway cloning vector set for high-hhroughput hunctional hnalysis of henes in planta. *Breakthrough Technologies*, ***133***(October), 462–469.

Cushman, J. J. C., & Bohnert, H. H. J. (1999). Crassulacean acid metabolism: molecular genetics. *Annual Review of Plant Biology*, *50*(1), 305–332.

Danker, T., Dreesen, B., Offermann, S., Horst, I., & Peterhänsel, C. (2008). Developmental information but not promoter activity controls the methylation state of histone H3 lysine 4 on two photosynthetic genes in maize. *The Plant Journal*, ***53***(3), 465-474.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., … Westhoff, P. (2008). The Gene for the P-Subunit of Glycine Decarboxylase from the C4 Species *Flaveria trinervia*: Analysis of Transcriptional Control in Transgenic *Flaveria bidentis* (C4) and Arabidopsis (C3). *Plant Physiology*, ***146***(4), 1773–1785.

Ernst, K., & Westhoff, P. (1997). The phospho*enol*pyruvate carboxylase (*ppc*) gene family of *Flaveria trinervia* (C4) and *F. pringlei* (C3): molecular characterization and expression analysis of this *ppc*B and *ppc*C genes. *Plant Molecular Biology*, ***34***, 247–443.

FAO IFAD UNICEF, W. & W. (2017). *The State of Food Security and Nutrition in the World*. *Fao*.

Gregorio, G. B., Senadhira, D., & Mendoza, R. D. (1997). *Screening rice for salinity tolerance* (Vol. 22, pp. 1-30). IRRI discussion paper series.

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M., & Westhoff, P. (2011). Evolution of C4 photosynthesis in the genus Flaveria: how many and which genes does it take to make C4? *The Plant Cell*, ***23***(June), 2087–2105.

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., &

Westhoff, P. (2004). *cis*-Regulatory Elements for Mesophyll-Specific Gene Expression in the C4 Plant *Flaveria trinervia* , the Promoter of the C_4 Phosphoenolpyruvate Carboxylase Gene. *The Plant Cell*, ***16***(May), 1077–1090.

Gowik, U., Schulze, S., Saladié, M., Rolland, V., Tanz, S. K., Westhoff, P., & Ludwig, M. (2017). A MEM1-like motif directs mesophyll cell-specific expression of the gene encoding the C_4 carbonic anhydrase in Flaveria. *Journal of Experimental Botany*, ***68***(2), 311–320.

Gowik, U., & Westhoff, P. (2011). The Path from C3 to C4 Photosynthesis. *Plant Physiology*, ***155***(1), 56–63.

Hartley, J. L., Temple, G. F., & Brasch, M. A. (2000). DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Research*, ***10***(11), 1788–1795.

Hatton, D., Sablowski, R., Yung, M.-H., Smith, C., Schuch, W., & Bevan, M. (1995). Two classes of cis sequences contribute to tissue-specific expression of a PAL2 promoter in transgenic tobacco. *The Plant Journal*, *6*(6), 859–876.

Hermans, J., & Westhoff, P. (1990). Analysis of Expression and Evolutionary Relationships of Phosphoenolpyruvate Carboxylase Genes in *Flaveria trinervia* (C4) and *F. pringlei* (C3). *Molecular & General Genetics*, ***224***(3), 459–468.

Heyndrickx, K. S., Van de Velde, J., Wang, C., Weigel, D., & Vandepoele, K. (2014). A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. *The Plant Cell*, tpc-114.

Hibberd, J. M., Sheehy, J. E., & Langdale, J. A. (2008). Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. *Current Opinion in Plant Biology*, ***11***(2), 228–231.

Hood, E. E., Gelvin, S. B., Melchers, L. S., & Hoekema, A. (1993). New Agrobacterium helper plasmids for gene transfer to plants. *Transgenic research*, ***2***(4), 208-218.

Hsieh, J., & Fire, A. (2000). Recognition and silencing of repeated DNA. *Annu. Rev. Genet.*, ***34***, 187–204.

Jefferson, R. A., Kavanagh, T. A., & Bevan, M. W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO journal*, ***6***(13), 3901-3907.

Kajala, K., Brown, N. J., Williams, B. P., Borrill, P., Taylor, L. E., & Hibberd, J. M. (2012). Multiple Arabidopsis genes primed for recruitment into C 4

photosynthesis. *Plant Journal*, ***69***(1), 47–56.

Kajala, K., Covshoff, S., Karki, S., Woodfield, H., Tolley, B. J., Dionora, M. J. A., … Quick, W. P. (2011). Strategies for engineering a two-celled C4photosynthetic pathway into rice. *Journal of Experimental Botany*, ***62***(9), 3001–3010.

Karki, S., Rizal, G., & Quick, W. P. (2013). Improvement of photosynthesis in rice (Oryza sativa L.) by inserting the C4 pathway. *Rice*, ***6***(1), 28.

Kawamura, T., Shigesada, K., Toh, H., Okumura, S., Yanagisawa, S., & Izui, K. (1992). Molecular evolution of phosphoenolpyruvate carboxylase for C4 photosynthesis in maize: comparison of its cDNA sequence with a newly isolated cDNA encoding an isozyme involved in the anaplerotic function. *The Journal of Biochemistry*, ***112***(1), 147-154.

Kellogg, E. A. (2001). Evolutionary history of the grasses. *Plant physiology*, ***125***(3), 1198-1205.

Kellogg, E. A. (1999). Phylogenetic aspects of the evolution of C4 photosynthesis. *C4 plant biology*, ***12***, 411-444.

Kirschner, S., Woodfield, H., Prusko, K., Koczor, M., Gowik, U., Hibberd, J. M., & Westhoff, P. (2018). Expression of SULTR2; 2 in the Arabidopsis bundle sheath and vein cells is mediated by a positive regulator. *Journal of experimental botany*.

Kosugi, S., Ohashi, Y., Nakajima, K., & Arai, Y. (1990). An improved assay for β-glucuronidase in transformed cells: Methanol almost completely suppresses a putative endogenous β-glucuronidase activity. *Plant Science*, ***70***(1), 133–140.

Ku, M. S., Agarie, S., Nomura, M., Fukayama, H., Tsuchida, H., Ono, K., ... & Matsuoka, M. (1999). High-level expression of maize phosphoenolpyruvate carboxylase in transgenic rice plants. *Nature biotechnology*, ***17***(1), 76.

Ku, M. S., Kano-Murakami, Y., & Matsuoka, M. (1996). Evolution and expression of C4 photosynthesis genes. *Plant physiology*, ***111***(4), 949.

Latzko, E., & Kelly, J. (1983). The multi-faceted function of phosphoenolpyruvate carboxylase in C3 plants. *Physiol Veg*, ***21***, 805–815.

Lazo, G. R., Stein, P. A., & Ludwig, R. A. (1991). A DNA transformation–competent Arabidopsis genomic library in Agrobacterium. *Nature Biotechnology*, ***9***(10), 963.

Lenhard, B., Sandelin, A., & Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, ***13***(4), 233.

Leyva, A., Liang, X., Pintor-Toro, J. A., Dixon, R. A., & Lamb, C. J. (1992). cis-element combinations determine phenylalanine ammonia-lyase gene tissue-specific expression patterns. *The Plant Cell*, ***4***(3), 263-271.

Mejía-Guerra, M. K., Li, W., Galeano, N. F., Vidal, M., Gray, J., Doseff, A. I., & Grotewold, E. (2015). Core promoter plasticity between maize tissues and genotypes contrasts with predominance of sharp transcription initiation sites. *The Plant Cell*, tpc-15.

Matsuoka, M., Kyozuka, J., Shimamoto, K., & Kano-Murakami, Y. (1994). The promoters of two carboxylases in a C4 plant (maize) direct cell-specific, light-regulated expression in a C3 plant (rice). *The Plant Journal*, ***6***(3), 311-319.

Matsuoka, M., Tada, Y., Fujimura, T., & Kano-Murakami, Y. (1993). Tissue-specific light-regulated expression directed by the promoter of a C4 gene, maize pyruvate, orthophosphate dikinase, in a C3 plant, rice. *Proceedings of the National Academy of Sciences*, ***90***(20), 9586-9590.

Melzer, E., & O'Leary, M. H. (1987). Anapleurotic CO_2 Fixation by Phosphoenolpyruvate Carboxylase in C3 Plants. *Plant Physiology*, *84*(1), 58–60.

Mitchell, P. L., & Sheehy, J. E. (2006). Supercharging rice photosynthesis to increase yield. *New Phytologist*, ***171***(4), 688–693.

Nomura, M., Higuchi, T., Ishida, Y., Ohta, S., Komari, T., Imaizumi, N., … Tajima, S. (2005). Differential Expression Pattern of C4 Bundle Sheath Expression Genes in Rice, a C3 Plant. *Plant and Cell Physiology*, ***46***(5), 754–761.

Offermann, S., Dreesen, B., Horst, I., Danker, T., Jaskiewicz, M., & Peterhansel, C. (2008). Developmental and environmental signals induce distinct histone acetylation profiles on distal and proximal promoter elements of the C4-Pepc gene in maize. *Genetics*.

Offermann, S., Danker, T., Dreymüller, D., Kalamajka, R., Töpsch, S., Weyand, K., & Peterhänsel, C. (2006). Illumination is necessary and sufficient to induce histone acetylation independent of transcriptional activity at the C4-specific phosphoenolpyruvate carboxylase promoter in maize. *Plant physiology*, ***141***(3), 1078-1088.

Sage, R. F. (2004). The evolution of C4 photosynthesis. *New Phytologist*, ***161***(2), 341–370.

Sage, R. F. (2016). A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of experimental botany*, ***67***(14), 4039-4056.

Schäffner, A. R., & Sheen, J. (1992). Maize C4 photosynthesis involves differential regulation of phospho*enol*pyruvate carboxylase genes. *The Plant Journal : For Cell and Molecular Biology*, ***2***(2), 221–232.

Schuler, M. L., Mantegazza, O., & Weber, A. P. M. (2016). Engineering C4 photosynthesis into C3 chassis in the synthetic biology age. *The Plant Journal : For Cell and Molecular Biology*, *87*(1), 51–65.

Schuller, K. A., Plaxton, W. C., & Turpin, D. H. (1990). Regulation of phosphoenolpyruvate carboxylase from the green alga *Selenastrum minutum*: properties associated with replenishment of tricarboxylic acid cycle intermediates during ammonium assimilation. *Plant Physiology*, ***93***(4), 1303-1311.

Sheehy, J. E., Ferrer, A. B., & Mitchell, P. L. (2008). Harnessing photosynthesis in tomorrow's world: humans, crop production and poverty alleviation. In *Photosynthesis. Energy from the sun* (pp. 1237-1242). Springer, Dordrecht.

Sheen, J. (1999). C4 Gene Expression. *Annual Review of Plant Physiology and Plant Molecular Biology*, ***50***, 187–217.

Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A., & Lukyanov, S. A. (1995). An improved PCR method for walking in uncloned genomic DNA. *Nucleic acids research*, ***23***(6), 1087.

Stockhaus, J., Schlue, U., Koczor, M., Chitty, J. A., Taylor, W. C., & Westhoff, P. (1997). The Promoter of the Gene Encoding the C4 Form of Phosphoenolpyruvate Carboxylase Directs Mesophyll-Specific Expression in Transgenic C4 Flaveria spp. *The Plant Cell*, ***9***(4), 479–489.

Studer, A. J., Schnable, J. C., Weissmann, S., Kolbe, A. R., McKain, M. R., Shao, Y., ... & Brutnell, T. P. (2016). The draft genome of the C 3 panicoid grass species Dichanthelium oligosanthes. *Genome biology*, ***17***(1), 223.

Toki, S., Hara, N., Ono, K., Onodera, H., Tagiri, A., Oka, S., & Tanaka, H. (2006). Early infection of scutellum tissue with Agrobacterium allows high-speed

transformation of rice. *Plant Journal*, ***47***(6), 969–976.

United Nations. (2015). World Population Prospects The 2015 Revision.

Weber, A. P. M., & von Caemmerer, S. (2010). Plastid transport and metabolism of C3 and C4 plants-comparative analysis and possible biotechnological exploitation. *Current Opinion in Plant Biology*, *13*(3), 257–265.

Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., … Westhoff, P. (2012). Regulation of the photorespiratory GLDPA gene in C(4) flaveria: an intricate interplay of transcriptional and posttranscriptional processes. *The Plant Cell*, ***24***(January), 137–151.

Yang, F., Li, W., Jiang, N., Yu, H., Morohashi, K., Ouma, W. Z., ... & Velazquez, R. A. (2017). A maize gene regulatory network for phenolic metabolism. *Molecular plant*, ***10***(3), 498-515.

Yanagisawa, S., & Sheen, J. (1998). Involvement of Maize Dof Zinc Finger Proteins in Tissue-Specific and Light-Regulated Gene Expression. *The Plant Cell*, ***10***(1), 75.

Zheng, X., Deng, W., Luo, K., Duan, H., Chen, Y., McAvoy, R., … Li, Y. (2007). The cauliflower mosaic virus (CaMV) 35S promoter sequence alters the level and patterns of activity of adjacent tissue- and organ-specific gene promoters. *Plant Cell Reports*, ***26***(8), 1195–1203.

Supplementary data

Supplementary Figure 1: Comparison of the proximal 500 bp of the 5′-flanking regions of *Ppc* genes from Panicoid C_4 grass *U. maxima* and C_3 grass *Dichanthelium oligosanthes*.

Supplementary table 1: List of oligonucleotides used for Genome walking

Supplementary table 2: List of oligonucleotides used to amplify promoter sequences

Supplementary table 3: Promoter-reporter constructs, and oligonucleotide combinations used for amplification

Supplementary table 4: List of the constructs transformed in the rice cultivars

Supplementary sequence data 1: 5′-flanking sequences of *Setaria viridis* C_4*Ppc* gene

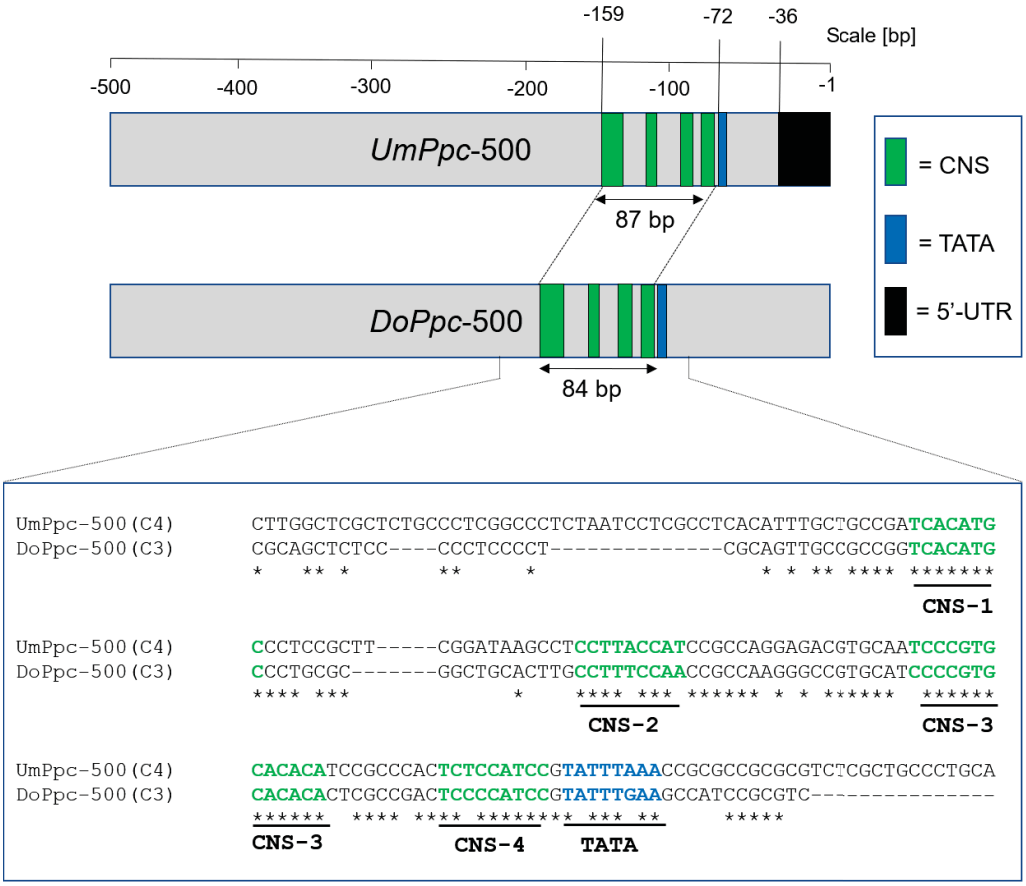
Supplementary sequence data 2: 5′-flanking sequences of *Panicum miliaceum* C_4*Ppc* gene

Supplementary sequence data 3: 5′-flanking sequences of *Urochloa maxima* C_4*Ppc* gene

Supplementary sequence data 4: 5′-flanking sequences of *Digitaria sanguinalis* C_4*Ppc* gene

Supplementary sequence data 5: Sequence comparison analysis of proximal 500 bp of the five-selected C_4 grass C_4*Ppc* 5′-flanking sequences

Supplementary Figure 1



Supplementary Figure 1: Comparison of the proximal 500 bp of the 5′-flanking regions of *Ppc* genes from Panicoid C_4 grass *U. maxima* and C_3 grass *Dichanthelium oligosanthes*. The schematic figure of top of the figure depicts the relative positions of the conserved nucleotide sequences (CNS), putative TATA box and the putative 5′-untranslated regions (5′-UTR). The CNSs are indicated in green, the putative TATA box in blue and the putative 5′-UTR in black rectangles. Relative length of the CNS harboring regions and the 5′-UTRs are indicated in bp below of each respective box. Nucleotide numbers refer to the translational start codon ATG (+1) and the positions are related to the black scale bar on top. TSS: putative transcription start site. The asterisk (*) indicates fully conserved base pairs

Supplementary Table 1: List of oligonucleotides used for Genome walking

Gene specific oligonucleotide	Sequence (5´ to 3´)
*S.italica*GW1	GAAGATGTCGAGGAAGCGGTCGATGA
*S.italica*GW2	GACGAGCTTGTCGTCCTCGGAGACCTT
*S.italica*GW3	GAGCTGCGCGTCGATGGAGTGGTGCTT
*S.italica*GW4	GCTAGATGGCTGTGGCGGTGGCACGA
*S.italica*GW5	GGCCTTGCGATGGGCTTGTTCCCAGTT
*S.italica*GW6	GCCCGGTGGCTCACTGACTCTGAAAAGC
*S.italica*GW7	GCATTACTACGATTGGCGCCGAGGGC
*S.italica*GW8	ATCTGTGGCAACGCCGTGGAGGATCAC
*S.italica*GW9	TGCCGTTGGCAACTCTGGAAGCACATG
*S.viridis*GW1	GCGTGCTTCTGAAGGAGCGACCTTCG
*S.viridis*GW2	TCCTCCTTGGACTTTCCGAGCTCTGT
*S.viridis*GW3	GCCTTGCGATGGGCTTGTTCCCAGTT
*S.viridis*GW4	ATTGGCGCCGAGGGCACTGATGTGTT
*S.viridis*GW5	GACGAAACCACCTGTGACCGAGCCCAT
*S.viridis*GW6	AGCAAGGCTGTGGCAATGCTGCTTTGT
*U.maxima*GW1	TGGAGCATGCAAGGTTTGGTGGGTTACG
*U.maxima*GW2	CAGACGGAGCTGCGCGTCAATGGAGT
*U.maxima*GW3	CCATCAGTAGAGCCTTGCCCCTGCTGT
*U.maxima*GW4	GTTTCGTAGCTGAGGTTGTGTGGCGCACTA
*U.maxima*GW5	ATCGGCAGGAGAGCCCCCTACCTGTT
*P.miliaceum*GW 1	CCGACAGCTCGTAGCAGTCCTGGACGAAT
*P.miliaceum*GW2	TACGACGTCCTCCTCATGGACCGCTTCC
*P.miliaceum*GW3	CAAGGTCTCCGAGGATGACAAGCTCGTCG
*P.miliaceum*GW4	GAGGTCTTCCAGCAAACCTATGCTGGCT
*D.sanguinalis*GW1	GAGGAGGGCATCGTACTCGACGAGGTT
*D.sanguinalis*GW2	CTCGGAGACCTTCCCAGGGGCAAGGA
*D.sanguinalis*GW3	GTTTCCAACCGGCTGCGATGATGCGT
*D.sanguinalis*GW4	CCCTCTTCCACTCTTCCTCTCTCCCTCTGT
*D.sanguinalis*GW5	GGCTTCTCCCCCATTGTTGCGGTCACT

Supplementary Table 2: **List of oligonucleotides used to amplify promoter sequences.**

Oligonucleotide name	**Sequence (5´to 3´), Gateway overhang sites are marked in *italic***
#*ZmPpc*-F	*GGGGACAAGTTTGTACAAAAAACAGGCT***TAATGGTGTTAGGACACGTG**
#*ZmPpc*-660-F	*GGGGACAAGTTTGTACAAAAAACAGGCT***CTTAGCCACAGCCGCCTCAC**
#*ZmPpc*-Rev	*GGGGACCACTTTCTACAAGAAAGCTGGGT***GGCGCGGCGGGAAGCTAAGC**
#*SvPpc*-F	*GGGGACAAGTTTGTACAAAAAACAGGCT***AAGTAGTTTGGGGGAATAA**
#*SvPpc*-Rev	*GGGGACCACTTTCTACAAGAAAGCTGGGT***GGCGTGGTGGGAAGCGAAGC**
#*PmPpc*-F	*GGGGACAAGTTTGTACAAAAAACAGGCT***TCGCCACACGTCAGCATCCT**
#*PmPpc*-Rev	*GGGGACCACTTTCTACAAGAAAGCTGGGT***GGCGTGCGCGAGCGGGAAG**
#*UmPpc*-F	*GGGGACAAGTTTGTACAAAAAACAGGCT***AATGCAGTCGGCGCCACCAG**
#*UmPpc*-Rev	*GGGGACCACTTTCTACAAGAAAGCTGGGT***GGCGTGTGCGTGCGTACGTG**
#01_*PmPpc*-Rev	*GGGGACCACTTTGTACAAGAAAGCTGGGTA***GGCGTGCGCGAGCGGG**
#04_*PmPpc*_500-F	*GGGGACAAGTTTGTACAAAAAAGCAGGCTTA***GACACCATTTTGCTCCAGTCG**
#05_*SvPpc*-Rev	*GGGGACCACTTTGTACAAGAAAGCTGGGTA***GGCGTGGTGGGAAGCGAA**
#08_*SvPpc*_500-F	*GGGGACAAGTTTGTACAAAAAAGCAGGCTTAA***CGGACACCTAGAGCGATAAT**
#09_*UmPpc*-Rev	*GGGGACCACTTTGTACAAGAAAGCTGGGTA***GGCGTGTGCGTGCGTAC**
#10_*UmPpc*_1500-F	*GGGGACAAGTTTGTACAAAAAAGCAGGCTTA***CAGGTCGCAATCGTTGCTAG**
#11_*UmPpc*_1000-F	*GGGGACAAGTTTGTACAAAAAAGCAGGCTTA***AATATCCTTATGTGTTCATAT TTCTTTG**
#12_*UmPpc*_500-F	*GGGGACAAGTTTGTACAAAAAAGCAGGCTTA***CTCCACACCAACCAACTTAGT**
#22_Fw_*UmPpc*-del_ 4	*GGGGACAAGTTTGTACAAAAAAGCAGGCTTA***CGCAGAGACAACCAGGACG**
#30_Um-blk-1-GO_F	*GGGGACAAGTTTGTACAAAAAAGCAGGCTTA***TGCCCTCGGCCCTCTAATC**

Supplementary Table 3: **Promoter-reporter constructs, and oligonucleotide combinations used for amplification**

Construct	Oligonucleotides combination for PCR amplification (5′and 3′)	Description of amplified/synthesized nucleotide sequences
ZmPpc:GUS	#*ZmPpc*-F #*ZmPpc*-R	Full length *ZmPpc* promoter sequences
SvPpc:GUS	#*SvPpc*-F #*SvPpc*-R	Full length *SvPpc* 5′- flanking sequences
PmPpc:GUS	#*PmPpc*-F #*PmPpc*-R	Full length *PmPpc* 5′-flanking sequences
UmPpc:GUS	#*UmPpc*-F #*UmPpc*-R	Full length *UmPpc* 5′-flanking sequences
UmPpc-1500:GUS	#10_*UmPpc*_1500-F, #09_*UmPpc*-Rev	Proximal 1500 bp of *UmPpc* 5′- flanking sequences.
UmPpc-1000:GUS	#11_*UmPpc*_1000-F, #09_*UmPpc*-Rev	Proximal 1000 bp of *UmPpc* 5′- flanking sequences
UmPpc-500:GUS	#12_*UmPpc*_500-F, #09_*UmPpc*-Rev	Proximal 500 bp of *UmPpc* 5′- flanking sequences
PmPpc-500:GUS	#01_*PmPpc*-Rev, #04_*PmPpc*_500-F	Proximal 500 bp of *PmPpc* 5′- flanking sequences
SvPpc-500:GUS	#05_*SvPpc*-Rev, #08_*SvPpc*_500-F	Proximal 500 bp of *SvPpc* 5′- flanking sequences
ZmPpc-660:GUS	#*ZmPpc*-660-F #*ZmPpc*-Rev	Proximal 660 bp of *ZmPpc* 5′- flanking sequences
UmPpc-500-del-4:GUS	#22_Fw_*UmPpc*-del_ 4, #09_*UmPpc*-Rev	Sub-fragment-4 of the *UmPpc*-500 was deleted
UmPpc-500-del-3:GUS	Synthesized by Biomatik	Sub-fragment-3 of the *UmPpc*-500 was deleted
UmPpc-500-del-2:GUS	Synthesized by Biomatik	Sub-fragment-2 of the *UmPpc*-500 was deleted
UmPpc-500-del-1:GUS	Synthesized by Biomatik	Sub-fragment-1 of the *UmPpc*-500 was deleted
UmPpc-500-region-1:GUS	#30_Um-blk-1-GO_Fw, #09_*UmPpc*-Rev	Sub-fragment-1 and the downstream sequences of the *UmPpc*-500

UmPpc-500-M35S:GUS	Synthesized by Biomatik	TATA and 5′-UTR segment of the *UmPpc*-500 was replaced with minimal 35S promoter
UmPpc-fGFP:GUS	Synthesized by Biomatik	Adjoining nucleotide of the CNS-1 and CNS-2, CNS-2 and CNS-3 were replaced with alien sequences.
SvPpc-500Δ92:GUS	Synthesized by Biomatik	CNSs harboring region of *SvPpc*-500 was deleted
*SvPpc*Δ92+*Um*87:GUS	Synthesized by Biomatik	CNSs harboring region of *SvPpc*-500 was replaced with the CNSs containing region of *UmPpc*

Supplementary Table 4: List of the constructs transformed in the rice cultivars.

Constructs in Kitaake cultivar	Constructs in Nipponbare cultivar
#*UmPpc*-1500:GUS	#*UmPpc*:GUS
#*UmPpc*-1000:GUS	#*SvPpc*:GUS
#*UmPpc*-500:GUS	#*PmPpc*:GUS
#*UmPpc*-500-del1:GUS	#*ZmPpc*:GUS
#*UmPpc*-500-del2:GUS	#*SvPpc*-500:GUS
#*UmPpc*-500-del3:GUS	#*PmPpc*-500:GUS
#*UmPpc*-500-del4:GUS	#*ZmPpc*-660:GUS
#*UmPpc*-region-1:GUS	
#*UmPpc*-M35S:GUS	
#*UmPpc*-fGFP:GUS	
#*SvPpc*-500Δ92:GUS	
#*SvPpc*Δ92+*Um*87:GUS	

Supplementary sequence data 1

5′-flanking sequences of *Setaria viridis C4Ppc* (5′to 3′):

TAAGTAGTTTGGGGGAATAAATCGAGCCTAAGTTTTATTTAGGGGTTAAGAGGACAGATAAAGTAGATT
GGTGATGGTAGTAAAATGGACATTTTCCTAGTTGTATCTGGTATGTACTATTTGGCTTAACATATATTA
AGTCTCAAGTAGTGCAAATTGGGATTCAAATACTAAATTCTCTTTACGTAGAGATGAGGAAGGAGAATT
ACAAAGCAGCATTGCCACAGCCTTGCTTCAGTTTTGCTCGATGCAAGCAGGGTGGAGTGGTGGACAGGC
ATGGACTAGCGGCACGGAAGACAACATCATGCATGTGCTTCCAGAGTTGCCAACGGCAAAAGAGCAGCA
ATGTGTGTAGTTCGGCCATGGGCTCGGTCACAGGTGGTTTCGTCTCCCAGAACAGGCAGGAGAGGGGGT
AGCTGGCTGAGCATCCATCCGTGATGAGCCTAAACCGTTGCGCAAACGGGCAGCTGGAGAGATGATGCG
AATATGTGATCCTCCACGGCGTTGCCACAGATGTCACGGATCATGCATGCATGTCGTCATTTAGCAGAT
GTTTGAGTCTGATACACCCTTCACACCAAGCAAATTATAACCAAGATAAGGATTCGAGGAAGCAGAAGG
AGGCAAGTGAGCAGTGACAAGTACGTCCACAACACAATCCAGACGAAAACGAACACATCAGTGCCCTCG
GCGCCAATCGTAGTAATGCTGCAACAAACAAAGTCCTGTGAAGACGGTCTCTATTAAATAGCCACACAT
GGAGCCGATGTATAATCCGAAAGATAATGTAATAATAGAGAAGCGTCATGTTTTTTCAAAAATAAAAAT
AAAAATAAACTGGGAACAAGCCCATCGCAAGGCCGGGAGCGAGCTTGCTTTTCAGAGTCAGTGAGCCAC
CGGGCGGGAGCTGACGGGTAGGGGAAGAGGAAGACAACATCGTGCATGTGCTTCAAGAGTTGCCAACGG
CAAAGGAGCAGCAGTGTGTGTAGTTCGGCTATGGGCTCGGTCACAGGTGGGTTTCGTCTCCCAGAACCA
GCAGGAGAGGGGGTTTGTGGCTGGCTGTGCATCCATCCGTGATGAGCCTAAACCGTTGCGCAAATGGAC
ACCTGGAGCGATAATGTGAATATGTGATCCTCCACGGCGTTGCCACAAATGTCACTGATCATCCATGTT
GTCCTATAGCACATGCTTGAGTCCAAGTCCCTCCACACCAAGCAAATTATAGCCAAGACAAGGATTCAA
GGAAGCAAGAAGGCGAGTGAGCAGTGACAAAGTATGTCCACAGCACAATCCGTACGAAAACGAGCGCAT
CAGTGCCCTCAGCGCCAATCGTAGTGTTTCCGCATCAAATAAGTCCTGTGAAGCCGGTCTCTATTAAAT
TGCCACGTATGAAGCCGATGTGTGGTCCGAAAGATAATATGCAATAATAGAGAAGCGTTATGTTTTCCT
GAAAAAAACCTGGGAACAAGCCTGTCGCAAGGCCGGGAGCCTCGCTTCCGTCTTGCTTTTTCACAGTCA
GTGAGCCACCGGGCGGGAGCTGACGGGTAGGGGAAGAGGAAGACAGCACCGTGCATGTGCTTCCAGAGT
TGCCAACGGCAAATGAGCAGCAGTGTGTGTAGTTCGGCCATGGGCTCGGTCACAGGTGGGTTTCGTCTC
CCAGAACCAGCAGGAGAGGGCGTATGTGGCTGGCTGTGCATCCATCATATCCGTGATGAGCCTAAACCG
TTGCGCAAACGGACACCTAGAGCGATAATGTATATGTGATCCTCCACGGCGTTGCCACAAATGTCACTG
ATCATGCATGTTGTCCTATATAGCACATGCTTGAGTCCAAGTCCCTCCACACCAAGCAAATTAATTATA
GCCAAGAGAAGGATTCGAGGAAGAAGAAGGCGATTGAGCAGTGACAAGTACGTCGACAGCACAACCCAG
ACGAAAACCAGGCAACGAGCACATCGCTCCGCAGCGCTGCCCCTCGGCCCTCCCCTTGCATTCGTGATT
CGCCGCCGG**TCACATGCCC**TCCGCTTGAAAAAAAAAAAGCCC**CCTTTCCA**TCCGCCATGAGGCGTGCAA
TCCCGTGCACACACTCGCCGA**CTCCCCATCC**GTTATTTTGTTTTTTTTAGAAGCCCCATCCGT**TATTTA**
AACCCACCCGCGCGCCTCGCTCGAC**CTCGTGCCACCGCCACAGCCATCTAGCTCCACTCCAGTCATCTA**
GCTTCGCTTCGCTTCCCACCACGCC
= 2233 base pair

5′ -UTR marked with bold black nucleotides, putative TATA box with blue, CNSs are in green.

The sequence data of *SvPpc* 5′-flanking sequences has been submitted to GenBank data libraries for publication and will be available soon under accession number MH675618.

Supplementary sequence data 2

5´-flanking sequences of *Panicum miliaceum C4Ppc* (5´to 3´):

TCGCCACACGTCAGCATCCTGGGCGGTGTCGCACCAGGTGCATCATGCGAATTTTCTTCCCTACTTGAA
AGTCAAGAGTTCTTTTTCTTAAAAAGCATCTATTTCAATTTCTCTCATATTCTAGCATTTTTTTGCTTA
AAACAAACAAAATATAATGCATCCCTCAAGTGTTAAAACTATGTCATATCATGCATGTTGGGAATGTTT
ATTGTGGTCTTGTGTGATGTGGTAGCCACAGCACCAACTTGGTTCCACATGTCCTTCTCTGTCACACAC
ACAATTACAAAATGTACCTCTCTATAGGAGTGCTAGTGTTGAACAATATCGCTGTTGGAATATAGGGAT
GGCCGCAAAAATGAAAAAAAACCTGTTTACTATGAATTATTTTTGAGATTCTCATCTTTTTGGTGTTCA
AATGAGAATGAAAATAGTATTTTAAAAACAAGAACGAAAATGATTATAAAGCAAGTAAACATAAACAAG
ATTTTGGGCTCTTTTAGTAATTTTTGAGATTTCCATATTAGACCGGTAATTACCGCCTAAAAATTACAA
CCAAATGTTTTCATGGTTAGATACACCACTCGCTCAAGTTCAACAACTAGGTGCCATCCTAAGAGCAAG
AGATGGTGAATTAGCACACTATTCCTGGGCTCCAACTAATCACTTGCTCAATGTAAACTTAACCTTGCA
TATATTGATGTGCTCTGAGGTTTATCTAGTGTGGCTAGAGATACAAACTTCGCACAAGTATTCATCCCA
TGGTAGTCGAATCAATGGATAGACTGTCATTCTAATCCATGTCGTTGCCATCAAGAATATGCTCCTAGT
CACTAAAACCCTCCCAATCAAGATTCTTGGATACCACCACCAAGATGGGCACCAAGCTGGATGGGCAAG
TTGCTCTTCAAGCCTCTCCTCCTATCACTTAGCGCCATCATTATTTTGGAGCATGAGCTGCAATAGCAA
GGGCCTATGTGAACCCGTGCAATCCTCTTGTCATCGCTTCATACACAAAACTAGAGAGTCAAGAAGGTG
AGCCCAAAGGCTAGCTCCAAGGCACCGGCACTCTATGCTTACAAGTGTGAACTATACTTTGATCTAACC
ACAATGAAGTGACATCTTATAGACAATTTATGTAGCCTAGACTAGAAATAATCCCCCTCTAATCTTGTG
CTAACTATCTTGGATCATGACTTTTACTTTGGTGGCTTGGTAGTTTTCTCAATAAGGTGTACCGTACCT
TCCTGTGGAGCAATAGGTAGTCATATTAGCTATAGAAATAGAGGTGCTCTTAGGCGAATGGCTAGGTCA
GATGCATGGGTGCGGTCAGTCCGCCTTCTGGCATGCGACAGCCTGCTCCTTTATGTAAATTCTACATTA
TTGACACCATTTTGCTCCAGTCATCTATGGCTTCCATTTCGATTTTCGAATGGCTTCGGCTCTTGCTCA
AACTGAACCACACCTGAGTCATCCTGACTCCTTGCAGGAGATGTGTGTGAGCTAAGCGTTGGTGACCTT
CCCTTCTCTGTGGTGGACATGAACCCAACTGGATGGCCCTGTATTGAGGACCTTCCATTGCAGCATTGG
AGAGAACACACAGCATGCAAATTCTGCCAAGGCGAGGAGCAGAAGAAGTCACGACCGACGGGTATGGTT
TTGTCAGGATCGCGTCAAGACACAACTGCAGCAGATGAAGACAAGGGAGGAGTGACAGGTATTCTTTTG
TCTCGATCGTGTCTACGGCCTGATCCAGACGAAAACGAGTACATCAGCCCTCAGCCCTGGCCCTTGCAG
AAGTTATATCAATAAACAGTTACTGCATTATCGATAATGTAATCTGAGTATCACAAAAGAATTTTATTC
CGGAGTTTTGAGTGGTTGCTGATAGCATTTCAGAACCAATAATTATGTATTGCTATAATCAACCCGCAA
TTTTGTATTCGAATAATAGCAAGTTAAGTGTTTGTGCCACACACATTGGCATAAAGTATCATGTTGTGA
AAGATAAAGTTCAGAATCAAATAATTAAGATTGAGAACATCAGTAATAAATTTATGCTTGCGGACTCAC
TTTCAAAAGGCCTATCACCTTGTATTTTTCGTAATCATGTAGCCAGCATAGGTTTGCTGGAAGACCTCT
GTTTCTGAACAAAAGAGAATCATTTATGAAACCAACTCTCGTAAAGTATTATGTGAATATCAGATAGAT
TTGGAATGTAATACTATAAATAATATGGTCAGGTAATTCATGATTGATCATCATGACACATTGCTCTAG
TGCGAAGTTTCATTTTGAAAGGATTTATGATGATTAACCTTAAAATTAAGGGGAGAATGTTGGATTAAT
CTCAAGGCTAACAATAATCCCGAGGTTAAAAGACCCATTAGGCCAAATTGCACCCTGATCGGGGCCCAT
AATCAATCCGAGGTTTCTGTCCGTGTCACCCAGTGTAATATTTAATAGAGAGGTTGGCCCCTTACACAA
ACTAATCTGTGAGGCTAAACCCTACTCCAACATTTCTAATCAGATCTAGCACGTGCGCTAAGCTGACGA
GAAGGCCACCGCCACCACTGCACCACCGTCACCAACGATATATCGTCCTCAAGCTCAGCGAACAATGTC
TCGTTCTTTAAATCTACGTGTTTTAGATCTAGAGATCTAGTGTTTACTCATGTTTATTTATGTTCTTAC
TCATGCAAATCTTACATTAATTCATCATCCCACAGTGTACCCTAAATATATAGCTCGCTATCTCAGAGA
CATGGATGCGGTCAGCCGGCCTTGTGGCGCGCGACTGCCTGCTCCTTTATGTAAATCGTACATTATTGA
CACCATTTTGCTCCAGTCGTTTTGATTTTCGAATGGCTTCGGATAGCTCTTGCTCGATCTGAACCACTC
CTGAGTCCTGAGTCCTTGCAGGGAGAGATGTGTGTGATCGAGCTAAGCGTGGGTGACCTTGAATTCTCT
TGTGTGTGGTGGACATGACCCCAACTGGATGGCCTTGTGCTGAGGACCTTCCATTGCAGTACGGCGAGT
CAAGACACGCCGCATGCGAGTTCTGCCAAGGCGAGGAGCAGAAGAAGGCAAGGTATGGTTTTGTCAGGA
TCGGATCGCTCCCCTGGCGGTGGCCGCCGG**TCACATGCCC**TCGCTCCGGTGGATAAGCCT**CCTTTCCA**A
CCGCCAAGAGGCCGCGCAT**TCCCGTGCACACA**CTGGCCCG**CTCCCCATCC**C**TATTTAAA**GCCACCCGCG

CTCTCGCCGTGC**ATCGCAACAGCGAGCTAGCACGGCGCGCCCTCCGCTCCAGCCACCTAGCTTGCCTTC CCGCTCGCGCACGCC**
= 3327 base pair

5′-UTR marked with bold black nucleotides, putative TATA box with blue, CNSs are in green.

The sequence data of *PmPpc* 5′-flanking sequences has been submitted to GenBank data libraries for publication and will be available soon under accession number MH675617.

Supplementary sequence data 3

5′-flanking sequences of *Urochloa maxima C4Ppc* (5′to 3′):

AATGCAGTCGGCGCCACCAGTTGACCATCGGGGCCAACTTCCATGATGTAGCCATCTTCTACCTCAACA
TAGAAGTAGACCTCTTCAGGTTCCTCCTCAACTTCCTCCACACGTGAAGCGGGAGGAGCAGACTGAGCG
GCAGGGATGCGAACCTGACGACCTCCGGTGCTTTTGCGAGCAGTTTGCTTAGTCCTTGCCATCTGACAG
AAGAGAGACCAAACTCAGAACAGTTTAAAACTAAGAAACCAGAGCTGACAAAGAAATAGAATAAAATAG
ATTTTTATGCAGAAATAAAGTGTTTGTGGAAGAAGAATTGGCTTGCTAAGAATTTGAGTCCTTAATTCG
ACCATTTCTAACTAGGCTAGCGTCCTGCAGTCAACACAGCTCTGATACCACTTCTGTCACAGCATCATT
GTAACTGTGGCGACAAGTCATTTTACTGTGATCCACCCTGTTCTTGCTACTCATCAACTGTGTCGATTA
GTCCTTTTGAATCGAGTGCTCGAACCCCTTTTGATTCGCTTCATTCATACTTGAAATTTCTGATTGCGT
TCTTATACTTTCTTGCTTGCGTTCTCGATTCGTTTGCAGGAAAAACCTTCTTGGCGAGGTCAATCAAGT
TCTCGCTTGGTTGATAACCAACGGAGCAGTAGTGTAGCAGGTTCGAACCGTGATGGTTTGAAGTCCCGA
TTGTTAACGTCGAGATCTCCACCAATCGAGTTTACCCTCTACCTCTCGGAATATCGGGCTTAGCTCTCA
TCATGGTATGGAGCGCATGAGGACTGGTACCTTTGGATGATTATCGACGGCGGGATGAGCTGATTGCAG
ACTGCGCTAAAAAAGCCCGACGTTAGACGATCTGAATCCGTCAATGCCACCTCAGTTCCTGCATGCAAC
CAAGAGTAGATTGACCAGCACACCAGCAAATAGACACCAAGCAAGAGTAGATTTTTTTATTTTGGAATA
CAAGCACGAGTAGATCGACAGGCAGAATGAGTAAAGGGCAAAAGGGACTTTTTAAATTTGTCTCATGCT
AGCATGGAGTACCACATCACTGCAAGGCCAAAAATAAGAGTTGGGGGACTAAATTGGCCGATATGAATG
TTGAGGGACTAACTTGACCCTTTGAGTTAAGTTGAGGGACCAATATGCCTATTTTGCCTAAATAATTCA
CTGATATTTCTATCATTTCCTGCTCATGAACTCAACAAGATCATTAGCTCTTTAATCTTGGTTTCATTC
AACTCACCAAAACCCACTAGGAGCCTAGATGCACTTTCAATGGCGCACCCGGTTCCTCCAACGGTATCC
TAGGTTAGGCTGGTTATCTTAATAAGTTGTCTCGTTAGAATCGTTCTTGTGAGGACGTGCAATGTAATA
ATGCTTGTTGGTTGAACAAATGCTTGATTTGGAGTTTGTAATGACTAAGATGTTTTGTGGTGTGATGGC
CGCAAATGCATTAATAAATAAATAAAGTTAAAGTCTGGCTAAGTTCAATCTTAGGTTTCTTGTTTTTCT
CTATCATGTGCCTTTATCAATTCATCATCCAGAGGTGCATTTTAGGTCGGTGCGTATGAACCGATGTGG
ATAAGGGTCGCTGTAGTAATGATGTCTTTACATAAACTTCCATTCATCTATCACATATTCGGTTGTGTC
TTCTCAAAAGGTGATCGGATTGGATCATCTGATCACTAAAGTTCTATCATAATTCACCTCTCAGAAGTC
CTATATCTTCATATATGCTTGGCACATAGCCATGTTCTCCACTTTTTAATTGAAGTCTGCTCAGAGCAA
GCACAAGTTCTTGCTTTTGCTAAGGGACTGGCTAAACACTAGGATGCTGAGAAGGAAAACCATGTTTTC
TGATGATTACTCCTGTGTGCTATGCTCTGCAGGTTGTGAAGAGACCTTGATGGCTTCATGCACCCTTTT
TTCAGATGTTTATTCAATCAAGCCTGCTGGGCATATTTGGGTATCATATGGGACATCACGTTACCCTAA
CTGGACATGATTTTGGTAGAGCCAGGTCGCAATCGTTGCTAGCTGGTGCATTTAATTTGGTGCCATCGC
AATAACTTCATTTTTGATAATGGACAGCGCTGACTAGCAAGATGGAAAGTCGAGTTCAGAGATGCAACT
CCTTTAGTTACGCTAAGAGGCAAACCTAGAGTGAAACTGGTGATAGATAATTGGATAAGTACTGTTCAA
TATTGCTCTGGTTTCTCTTTCTTTAAGCTTGAAGCTTTGTAGTCCCCTTGTACATAAAACATTTATATT

TATATAAGAAAACAGGTAGGGGGCTCTCCTGCCGATTTTCTAAAAAAAAAAGAAACAAGAGTACTCAAA
TCCAGAAAAATAGGGTATGACATTCATATGGTAAGAAATCTCTATAAAAAATGAATAGTGCGCCACACA
ACCTCAGCTACGAAACATCCGCAAGAATAAAAATCTGTAGTTCAATGGTGAATCGGTAATGTTCTGACA
ACAAAATGATGGCACTCAGTCGTTTATAATTCTCCGATGAATATCCTTATGTGTTCATATTTCTTTGCA
TTGGACCATCCGACATTTTGCTCTACTCTGGTGAGAACAATTTTCCTAGGATTTGTCCAATTCAATCAA
AATTTTGTCTCAACTTTGGTGGCTTTATTAAAAAATAAAAAACAGCAGGGGCAAGGCTCTACCTCTACT
GATGGGAAAAACAGGAACAACGAAATTTCAAGGCTGCAAGCCAAGCTCTTCCATTAATCAGTGCTTCAC
TCAGAGACATGGATGCGATCATTCTTATGGTGCGTGACTGCCTTATCCTTTTTGTGAATTGTACATTGC
AACCATTTTGCTCCAGTCGTGAAGGGCTTCCACTTTGTGTCGAATGTCTTTGCCACTTGCTCGAACTGC
AACCACTCGTCTCAGGAGATGGGTGAGCAAAAGCGTCGGTGACCTTCTTTGTGGGGAACATTAACCCAA
CTGGCCTCACTTGTGGTCGTGGACTGTCGATTACAGTATGCGAGTCAAGGCACCGTCCAGGCCTCCACA
CCAACCAAATTAGTCGAGAGTGGAGACAAGAAGCATAAGAAGGCAAGTGAGCTGTGCCAAGCATGTCTC
CGTCTGGGAGTCTGGACTGTGTGCCGCAGAAACAACCAGGACGAAAACGAGCACATCACGGCCTTTGTG
GTTGTGGACAATATGAGCCAATACACTCTCTACACGTAGCAAATCAAGCCAAGGCAAGGCGCAGAAGAA
GGCAAAGTGAGCAGCGACGCCGCTTCGTCCTGACTGTGCCGCGGACACAATTACTCCGAAAACGAGCAC
ATCAGGCCTTGGCTCGCTCTGCCCTCGGCCCTCTAATCCTCGCCTCACATTTGCTGCCGA**TCACATGCC**
CTCCGCTTCGGATAAGCCT**CCTTACCA**TCCGCCAGGAGACGTGCAA**TCCCGTGCACACA**TCCGCCCA**CT**
CTCCATCCG**TATTTAAA**CCGCGCCGCGCGTCTCGCTGCCCTGC**ATCACAGGAACCTCACAGCACGTACG**
CACGCACACGCC
=3531 bp

5′-UTR marked with bold black nucleotides, putative TATA box with blue, CNSs are in green.

The sequence data of *UmPpc* 5′-flanking sequences has been submitted to GenBank data libraries for publication and will be available soon under accession number MH675616.

Supplementary sequence data 4

5′-flanking sequences of *Digitaria sanguinalis C4Ppc* (5′to 3′):

AAATTCAAATCTAACTTTTATATTCAAATAAAGTCCAAATAAATTCCAAATGAATCCCAAATGGAACTC
CAAATGTCACACTGAATTATATAGGTCCTTAGCTCAATAAAATATTACTTGAAATTTTGAATTCAAGGT
GAGTTCACTTGGAAATTCAAAGGTTTGCACAATATTCATAGAAAAATTTTAAAAATTTCACCTTTTTGA
AATTAGGGATATTACACCCAGGCTCTCTCGTCGCCAGCTCACCACTACCACCAGGAACGCACACATGTT
GCACGTAGGCAACCAGGCCAGCATCGGTGCACACATGTTGCACGTCGCTCCGCCGACCACGTCGCAACA
TCGCAGAGTCGCCAGCTTAGCCTCTCACGGATCCGTGAGGTGGATAGGATAGGCATCATTGCCTGGATG
GTGCATTGGAGCTGCCAACTCCCTCCTTGGCTCACCTACTTGTCCATGGCGGCATGCCCTTTTGCATGC
CTAATGCTAAGCACCATCAGTCACAGGAAGAATAGGAGCTTTTTTTTTGTAGCCATGCCATCCGGAGTA
CCTGTTTTCTCTTGCTGTCAACATTTCTAACACTGATGACCCGCCATGTAAACTCAGCTAAATGTTCGA
AGTGGTTACATATGTCGTTGTTCTTGATCGTAATTGATATTTTTTTGATTGTAATCAACATGTTACCAA
TCGAATTATCCTATTCCAGATATCCCGTAATATCGCTTTTGTTTTCCCGTCTGACTTCCCCTTTCCCGC
TCCCATTCCCATCCAGAAAACATAGCAGCGGGAATGGTTCAGGTGTTTCCTGACCCTTCCCGACTGCGT
TCATCCCTACTTCTCCACCCTCTCTATTCCCAATGAGCTTCGACCCATTCTCGGCTGGGAAACGAAGCT
CTCCATCCGCCATTGCTGCCTACGGGGCTGAGATCCGCCCGTAGGCGAACTCGCATCACCGGCGCCCAT
TTTCCTTGCCAAGACTCAAATCGAATATAAAAGGCCTCCTAAAGCTCTTGCCCTCTACGTTCCACCATG
TTCTATTGGTTTCCGCTCTCATCCATGGAACTCAAAGTCGGCTGGAATCCTGTCCTTGCTGTCGGCCAT
GCTCAACAATTGCTCGTCACGCACATCACCCACCGCTGAACTCACCTCACTGCGTAGATGTTGTAGTTG
CAACTCCTTGCTATCATACGGCTCGGTGCCAGCGGAATACACTGGTCAACTGGCTGCCTGCCACGATTT
TCACCAGCAAAATGAGAGAGAAAAGAGAAAATATGAGTTTTCATTTGGGCCCCACTTCTATGTGTTGTC

CATGTGAGTGAAATCACCCTTCAAAACAGCTGGATGGCCAATGGTTGTGGTAGTTTGATTGTCAATAAT
ATCCGGTTTTGGAGTTGGCCAAAATCGAACTCAGTTGATAGTTTGGATGGCTAAAATGGACATTTTTCC
AAATCTACAATACATAATTGATTGCTTACAACTCATGCAAAACAGAGAGCATTACAAAACATTCCAGGT
ATGCAAAAACATTCCATATCTATGTACAGACAGACCAAAAATTAACGAGGGTAATTCTTGCTCTAGCGC
CACAACTGTATCCATTAACCACGTCGCCAGCAGAGAGCATTAAAAACAATCCGTATATGCATCCAAGGA
CCAGCAGAGAACTGGCCGTTTGAGGCAGGATTTTCAAGAACTGGCTGTTCCAAGCGGGATTCATGGGAT
TTGATATCAAAAGCCTATCATCTGATGAACTGGCCAGAATCTCCCGTTCTGGAAATCAGGAGGGAAGGG
GGCAGCAACGGCTAGTATCCACGCGATGTATAAGTGACTCGCCCAGGAGAGAAAGAGGCAGCAGAGAGG
AATTCTGACCTGAAATCCCCAGGCGCCGTGCTGCTCAATCCCAAGCCCGCCATCGTCTCTTGGCCGCCG
ACTCGTCACCGTGTCCAAGTCTCGAGACGCCGCCGTGTCAAGCTGCAGCTGCCCAGCACTGCGCAGGTA
ACAGAGATGGACCTGGCTATCGATCTGGCTTGATTCGTCCGTGGAGTACATGTGCGTCCATGTGGTTCG
TTCGTGGAAGACACTCGGCACACTTGATTCCACTACTACAATAGCGCCATTACAGCCGGTTCGCAACCC
CCCTTCGCAGCCAGTTTTGCGAACCGTTTGATTTCAAGCGGTAGCGTTGATCCCCTCATCGCAGCCGGT
TATAAACCGGCTGTGACACAAGTCATCACAGCCGGTTGGTAACACGCACCGGCTGTGTTGACAGCCACA
ACACAGCCGGTTGGTGACATACACCGGCTGTGTTGTGTCAGTCAACACAGTCAGTTGGTAATACCAACC
GGCTGTGTTGAGCAACACAACACTGCCGGTTCGTAACATGCAAATAGCACCTACCACCAAGAACATCAA
ACCCATCAATTTCAATTCAAAAATGCAAATTAGAGGCACTAATCAAGCAAAGGATGTGATACATTGCAA
GAAACTAACCTTAACACAAGTTGGTGAAGTCGAATCCTCAAAGGAATGGAGTTTCCATGAGCTTCATGG
TGAAAAGTGACCGCAACAATGGGGGAGAAGCCAAGAGCCGTCGGGTTGGAGAAGAAAACAGAGGGAGAG
AGGAAGAGTGGAAGAGGGTTTTGTACTGTACGCATCATCGCAGCCGGTTGGAAACACCAACCGTTTGTG
ATATGTCCATTCACAGCCGGTTGGTGTCACAAACCGGCTGCGTTGCCCCGGCTGATCATCGAAGCCGGT
TGGTAACACCAACTGGCTGCGATGAGTTTCATCACAGCCGGTTCCAACGGTTACGAGGGCCACCCACTT
TGAAACCGGCTGTGATGTCTGTGTGCAGCCGGTTGGCCCAAAACCGGCTGCGAAGGGGGCGTTTACAAT
GACCGATTCTGTAGTAGTGTATTAATTTTAATTAGCCCGTGCAGTAGCACGAGTTGATAGGCTAGTGTT
TATTAATTACATATACCAACATGCACCATCTTCTCAGTGCTTGGTCTATCTCTGAGGGTTTTCTGTTTT
TTGCCTAATTCATCTAGTCATGACACCACCCATACCTGTCATCGGGGAGGGTTTGGGCGATCCGTACGC
CCCACTAATTCCCCGTACCCATACTGAATGAAATTATCTGTACCTAATTAAAATGGATAATGCCTGATG
GGGTGGATATATCCAGGTGAGGAAGTCCACGACCACGAGCAGCAGCTCCATGCGGGTAGCAGTGCTTGG
AACTGTGTTGCGCGACATGCCAACAGGAGCTTAGCGAGATCGGGTGGTACGGGGAAATTGAGATCCAGA
GACTGTATTATCCTACTATATACCTATATCGAAATTGTAATGGGTATAATTGTTCCCCCCATTCCTGAC
CCTAATACCATGGATTGATGGATATGAATTTAGGGGTGGGCATTGGATCAGCCCCGGCAGGCATAGCAC
CACCATCCATCTTCGATAGATTTGTCCTTATTAAACATATGGTCTCTAATCTTGAACACGAAAGATTGC
AGCATATGCGGATCATCAGGCACTGCCCTCCACACCAAACAAAACAGCAAATGCAAGAAGCCGAAGATA
GCAGGCGAGCAGCGACGAGCACGTCCACAACACCGCAAGATCCCGACGGAAATAAGCACATCCATCCGC
ATCTCGGTCCGATATGAAAAAAATGACAGCATTAATTAGGGGGCATATATAATTCAGACGATACTCCCC
TCCACATCAAACAAATCATCCAACGCAACGACACAAAAACACCGCACAAGGTGGCAAGCAAGCAGTGAC
GAGCACGTCCACAGCACAGACCACAAAAGATTGCAGCAATGAGCACAAAAGATTGCGGTATATATGCGG
CTCAGATAATACCCCTCTACACCAAACAAATTAGACAATGCTAGAAGCAGAACAAGGGCAGGCTAGCAG
TGATTAGCACGTCCACAGCACAGCACGATCCCGACGAAAACGAGCACATCCGTCAACATCTCAACCCGT
CGGTACTCACTGGTCCTACCCTGGCAGTATCTCACAG**TCACATGCCC**ACCGCTCGAAAAAAAAGCTT**CC**
TTTCCATCCACCAGGAGGCGTGCAA**TCCTGTGCACACA**CTCGCCGA**CTCCCCATCC**G**TATTTATA**GCCC
TCCACGGGACTTGCTCTCGT**GCATCTTTGCAACAGCACACCACTCTAGCAAGCCGTCCTCGAGCTTGCC**
TTCCCGCCACGCC

=4222 base pair

5′-UTR marked with bold black nucleotides, putative TATA box with blue, CNSs are in green.

The sequence data of *DsPpc* 5′-flanking sequences has been submitted to GenBank data libraries for publication and will be available soon under accession number MH675615.

Supplementary sequence data 5

Sequence comparison analysis of proximal 500 bp of the five-selected grass *C4Ppc* 5´-flanking sequences

```
CLUSTAL O(1.2.4) multiple sequence alignment (5´to 3´)

UmPpc-500       CTCCACACCAACCAACTTAGTCGAGAGTGGAGACAAGAAGCATAAGAAGGCAAGTGAGCT
SvPpc-500       -ACGGACACCTAGAGCGATAATGTATATGTGATCCTCCACGGCGTTGCCACAAATGTCAC
SiPpc-500       --CGGACACCTAGAGCGATAATGTATATGTGATCCTCCACGGCGTTGCCACAAATGTCAC
PmPpc-500       -----GACACCATTTTGCTCCAGTCGT--TTTGATTTTCGAATGGCTTCGGATAGCTCTT
DsPpc-500       ---CCCCTCCACATCAAACAAATCATCCAACGCAACGACACAAAAACACCGCACAAGGTG
ZmPpc-500       ---TCCAGCATTAACTGCTAAGGGACGCCCTCTCCACATCCTGCTACCCAATTAGCCAAC


UmPpc-500       GTGCCAAGCATGTCTCCGTCTGGGAGTCTGGACTGTGTGCCGCAGAGACAACCAGGACGG
SvPpc-500       TGATCATGCATGTTGTCCTATATAGCACATGCTTGAGTCCAAGTCCCTC---------CA
SiPpc-500       TGATCATGCATGTTGTCCTATATAGCACATGCTTGAGTCCAAGTCCCTC---------CA
PmPpc-500       GC------TCGATCTGAACCACTCCTGAGTCCTGA-GTCCTTGCAGGGAGAGATGTGTGT
DsPpc-500       GCAAGCAAGCAGTGACGAGCACGTCCACAGCACAGACCACACAAAAGATTGCAGCAATGAGC
ZmPpc-500       GGAATAACACAAGAAGGCAGGTGAGCAGTGACAAAGCACGTCAACAGCACCGAGCCAAGC


UmPpc-500       AAACGAGCACATCACGGCCTTTGTGGTTGTGGACAATATGAGCCAATACA-CTCTCTACA
SvPpc-500       CACCAAGCAAATTAATTATAGCCAAGAGAAGGATTCG-----AGGAAGA-----AGAAGG
SiPpc-500       CACCAAGCAAATTAATTATAGCCAAGAGAAGGATTCG-----AGGAAGA-----AGAAGG
PmPpc-500       GATCGAGCTAAGCGTGGGTGACCTTGAATTCTCTTGTGTGTGGTGGACATGACCCCAACT
DsPpc-500       ACAAAAGATT---GCGGTATATAT----GCGGCTCA-----GATAATAC--CCCTCTACA
ZmPpc-500       CAAAAAGGAG---CAAGGAGGAGC----AAGCCCAA-----GCCGCAGC--CGCAGCTCT
                     **

UmPpc-500       CGTAGCAAATCAAGCCAAGGGAAGGCGCAGAAGAAGGCAAAGTGAGCAGCGACGCCGCTT
SvPpc-500       CGATTGAGCAGTGACAAGTA-CGTCGACAGCACAACCCAGACGAAAACCAG-------GC
SiPpc-500       CGATTGAGCAGTGACAAGTA-CGTCGACAGCACAACCCAGACGAAAACCAG-------GC
PmPpc-500       GGATGGCCTTGTGCTGAGGACCTTCCATTGCAGTACGGCGAGTCAAGACAC--------GC
DsPpc-500       CCAAACA-AATTAGACAATGCTAGAAGCAGAACAAGGGCAGGCTAGCAGTGA------TT
ZmPpc-500       CCAGGTC-CCCTTGCGATTGCCGCCAGCAGTAGCAGACACCCCTCTCCACA-------TC
                        *           * *  *

UmPpc-500       CGTCCTGACTGCCGCGGACACAATCCGGACGAAAACGAGCACATCAGGCCTTGGCTCGCT
SvPpc-500       AACGAGCA----------CATCGCTCCGCAGCGCTGCCCCTCGGC---------------
SiPpc-500       AACGAGCA----------CATCGCTCCGCAGCGCTGCCCCTCGGC---------------
PmPpc-500       CGCATGCG----------AGTTCTGCCAAGGCGAGGAGCAGAAGAAGGCAAGGTATGGTT
DsPpc-500       AGCACGTCCACAGCACAGCACGATCCCGACGAAAACGAGCACATCCGTCAACATCTCAAC
ZmPpc-500       CCCTCCGGCCGCTAACAGCAGCAAGCCAAGCCAAAAAGGAGCCTCAGC----------CG
                                      *
UmPpc-500       CTGCCCTCGGCCCTCTAATCCTCGCCTCACATTTGCTACCGATCACATGCCCTCCGCTT-
SvPpc-500       ------------CCTCCCCTTGCATTCGTGATTCGCCGCCGGTCACATGCCCTCCGCTTG
SiPpc-500       ------------CCTCCCCTTGCATTCGTGATTCGCCGCCGGTCACATGCCCTCCGCTTG
PmPpc-500       TTG---TCAGGATCGGATCGCTCCCCTGGCGGTGGCCGCCGGTCACATGCCCTCGCTCCG
DsPpc-500       CCGTC-GGTACTCACTGGTCCTACCCTGGCAGTATCTCACAGTCACATGCCCACCGCTCG
ZmPpc-500       C--------------AGCCGGTTCCGTTGCGGTTACCGCCGATCACATGCCCAAGGC---
                                        *  *   *  **********

UmPpc-500       -----CGGATAAGCCTCCTTACCATCCG-CCAGGAGACGTGCAA-TCCCGTG------CAC
SvPpc-500       AAAAAAA-AAAAGCCCCCTTTCCATCCG-CCATGAGGCGTGCAA-TCCCGTG------CAC
SiPpc-500       AAAAAAAAAAAAAGCCCCCTTTCCATCCG-CCATGAGGCGTGCAA-TCCCGTG------CAC
PmPpc-500       ----GTGGATAAGCCTCCTTTCCAACCGCCAA-GAGGCCGCGCATTCCCGTG------CAC
DsPpc-500       ----AAAAAAAAGCTTCCTTTCCATCCA-CCAGGAGGCGTGCAA-TCCTGTG------CAC
ZmPpc-500       ------------CGCGCCTTTCCGAACG-CCGAGGGCCGCCCGT-TCCCGTGCACAGCCAC
                          **** **   *  *   * * *       *** ***      ***

UmPpc-500       ACATCC-----------GCCCACTCTCCATCCG---------------------------
SvPpc-500       ACACTC-----------GCCGACTCCCCATCCGTTATTTTGTTTTTTTTAGAAGCCCCAT
SiPpc-500       ACACTC-----------GCCGACTCCCCATCCGTTATTTTGTTTTTTTTAGAAGCCCCAT
PmPpc-500       ACACTG-----------GCCCGCTCCCCATCCC---------------------------
DsPpc-500       ACACTC-----------GCCGACTCCCCATCCG---------------------------
ZmPpc-500       ACACACCCGCCCGCCAACGACTCCCCATCCC-----------------------GAT
                ***               *  *** ******
```

```
UmPpc-500       ----TATTTAAACCGCGCCGCGCGTCTCGCTG------CCCTGCATCA-------------
SvPpc-500       CCGTTATTTAAACCCACCCGCGCGCCTCGCTCGACCT-CGTGCCA---CCGCCACAGCCAT
SiPpc-500       CCGTTATTTAAACCCACCCGCGCGCCTCGCTCGACCT-CGTGCCA---CCGCCACAGCCAT
PmPpc-500       ----TATTTAAAGCCACCCGCGCTC-TCGCCGTGCAT-CGCAACAGCGAGCTAGCACGGCG
DsPpc-500       ----TATTTATAGCCCTCCACGGGACTTGCT-------CTCGTGCATCTTTGCAACAGC--
ZmPpc-500       ----TATTTGAACCCACCCGCGCAC-T-GCATTCACCAATCGCATCGCAGCAGCACGAGCA
                    *****  *     **  **   * **

UmPpc-500       CAGGAACCTCACAGCACGTACGCACGCAC--------------------ACGCC***
SvPpc-500       CTAGCTCCACTCCAGTCATCTAGCTTCGCTTCGCTT-------CCCACCACGCC***
SiPpc-500       CTAGCTCCACTCCAGTCATCTAGCTTCGCTTCGCTT-------CCCACCACGCC***
PmPpc-500       CGCCCTCCGCTCCAGCCACCTAGCTTGCCTTCCCGCTCGCGC-------ACGCC***
DsPpc-500       ACACCACTCTAGCAAGCCGTCCTCGAGCTTGCCTTCCCG--------CCACGCC***
ZmPpc-500       GCACGCCGTGCCGCTCCAACCATCTCGCTTCCGTGCTTAGCTTCCCGCCGCGCC***
                                                                ****
```

The authors' contribution

SDG wrote the manuscript and performed all the experiments except those listed below

ML performed the 5´-RACE analysis and the vectorette PCR (genome walking)

SS prepared the *SvPpc*:GUS, *PmPpc*:GUS, and *UmPpc*:GUS constructs and performed GUS histo-chemical and quantitative analysis of the three respective constructs.

SK & **PQ** helped with the transformation of the *ZmPpc*:GUS, *SvPpc*:GUS, *PmPpc*:GUS, *UmPpc*:GUS, *ZmPpc*-660:GUS, *SvPpc*-500:GUS, and *PmPpc*-500:GUS constructs in the Nipponbare rice cultivar.

JE prepared and analyzed the GUS histo-chemical and quantitative analysis of *ZmPpc*:GUS and *ZmPpc*-660:GUS constructs

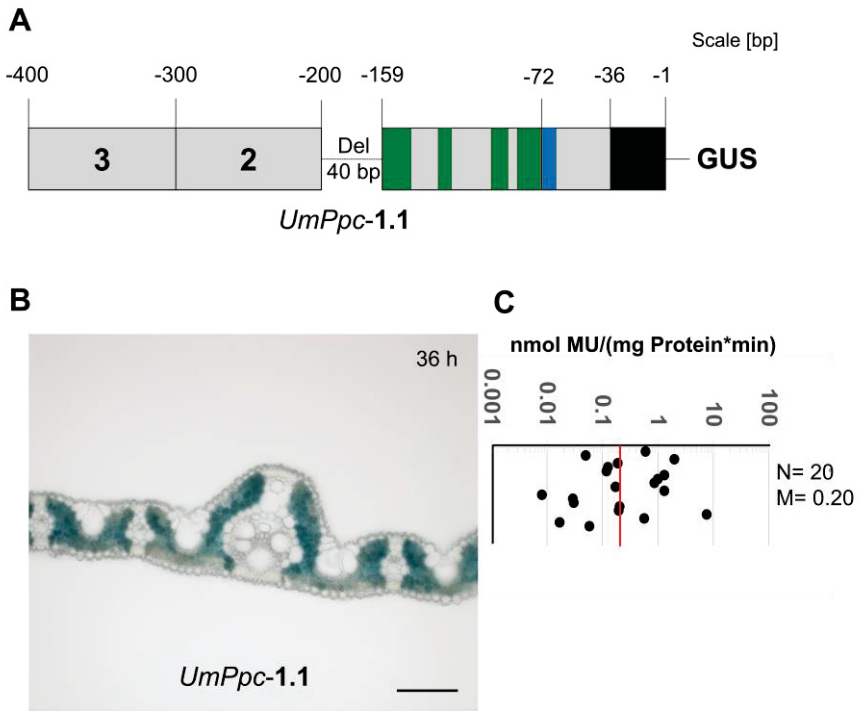
MS helped with the rice transformation of *SvPpc*-500Δ92:GUS, *SvPpc*Δ92+*Um*87:GUS, and *UmPpc*-fGFP:GUS constructs.

PW & **UG** will participate in drafting of the manuscript

Addendum

Deletion of the proximal 40 bp from sub-fragment-1 of *UmPpc* but keeping the CNSs intact reduces the promoter activity

In the 5´-flanking sequences of *UmPpc*, the region containing the four putative CNSs is located in sub-fragment-1 at the position from -159 bp to -73 bp. In order to know whether it is possible to do further narrow down the essential sequence in sub-fragment-1, we deleted 40 bp nucleotides from the 5´ ends of sub-fragment-1 (-200 bp to -160 bp) while maintaining the CNSs and named the truncated sub-fragment-1 as sub-fragment-1.1. To investigate the functionality of sub-fragment-1.1, we fused this region with its upstream sequence (sub-fragment-2 and sub-fragment-3 segments of the *UmPpc* promoter) to the GUS reporter gene. The subsequent construct *UmPpc*-1.1:GUS was analyzed in Kitaake rice (Addendum figure 1).



Addendum figure 1: Functional analysis of 5´-deletion of 40 bp from sub-fragment-1.

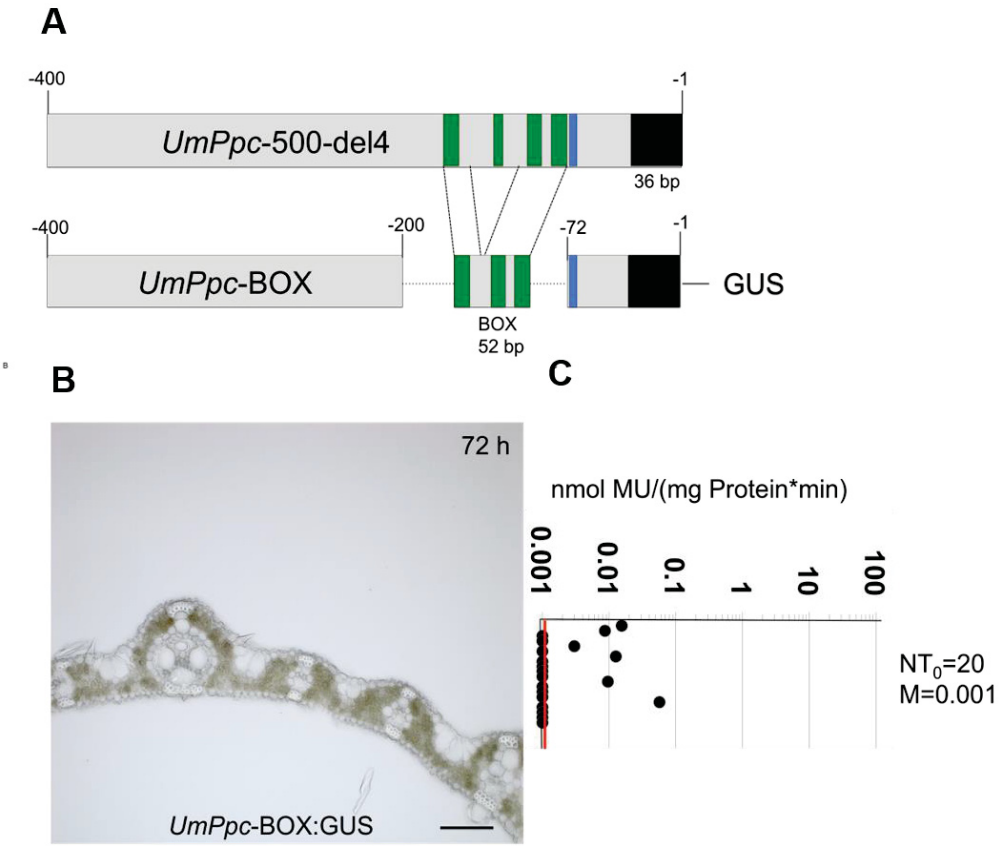
(A) Schematic presentation of the *UmPpc*-1.1:GUS construct. **(B)** Histochemical localization of GUS activity in a leaf cross section of transgenic rice transformed with the *UmPpc*-1.1:GUS construct. Incubation period in GUS staining buffer was 36 hours. Scale bar: 50 µm. **(C)** GUS activities in leaves of transgenic rice plants. GUS activities are expressed in nmol of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number (NT_0) of transgenic plants analyzed and the median (M) values of GUS activity are represented at the top of column. The position of median value is marked in red line.

After analyzing leaf cross sections of 20 transgenic plants of the *UmPpc*-1.1:GUS construct, we found only eight transgenic lines showing GUS staining upon incubation in GUS staining buffer for 18 to 72 hours (Addendum figure 1). Most of these plants (seven out of eight plants) required a longer staining period (48 to 72 hours) to detect the blue staining in the M cells of rice leaves indicating reduced activity of the promoter. The result from the histochemical assay could be correlated with the result of the GUS-fluorometrical quantitative assay (Addendum figure 1C). The GUS fluorometric analysis also showed that only eight transgenic lines of *UmPpc*-1.1:GUS construct have the activity value of histochemical detection level threshold 0.4 nmol MU/(mg protein * min). The rest of the 12 plants is below this threshold level; thereby they did not show any blue staining in the histochemical analysis. The substantial decline in GUS activity for *UmPpc*-1.1:GUS construct suggests that there might be one or more positive regulator(s) within this proximal 40 bp of sub-fragment-1. Since sub-fragment-1 alone itself cannot drive GUS expression in transgenic rice (Figure 5), this indicated that the positive regulatory element(s) could be overlapping the border between sub-fragment-1 and sub-fragment-2.

Deletion of CNS-2 along with the adjoining nucleotides causes complete loss of activity

The data presented in the manuscript results part clearly demonstrated the pronounced importance of the CNSs harboring region for the transcriptional activity of the C_4 grass *C_4Ppc* promoters. The question is, are all four CNSs important for functioning? Do they work additively as a common *cis* element(s) in the C_4 grass *Ppc* promoters? If all CNSs are the *cis* element(s), then one could speculate that deletion of any of the CNSs should produce a substantial difference in the expression of the reporter gene. To address these questions we randomly deleted the CNS-2 along with

some additional adjacent nucleotides (total of 35 bp) from the *UmPpc*-1.1:GUS construct resulting in construct *UmPpc*-BOX:GUS (Addendum figure 2). Since the adjoining nucleotides of the CNS-2 did not show any impact upon replacing with a foreign nucleotide sequence (Figure 7) we deleted most of those sequences. The other three CNSs (CNS-1, CNS-3, and CNS-4) were kept intact in this construct, and we examined its expression pattern in the stably transformed Kitaake rice cultivar (Addendum figure 2).



Addendum figure 2: Deletion analysis of the CNS-2 along with the adjacent sequences.

(A) Schematic presentation of *UmPpc*-BOX:GUS construct. The CNS-2 and the adjoining nucleotides (total of 35 bp) were deleted from the *UmPpc*-1.1:GUS construct. **(B)** shows the GUS histochemical results of cross section of a leaf blade of a rice plant transformed with the *UmPpc*-BOX:GUS construct. Incubation period in GUS staining buffer was 72 hours. Scale bar: 50 µm. **(C)** represents MUG activity assay data of *UmPpc*-BOX:GUS transformants leaf extract. GUS activities are expressed in nmol of the reaction product 4-methylumbelliferone (MU) per mg of protein per minute. The number (NT_0) of transgenic plants analyzed and the median (M) values of GUS activity

for the construct are represented at the top of the column. The position of median value is marked in red line

The data of the GUS histochemical analysis and the MUG quantitative assay showed that the removal of the respective 35 bp along with the CNS-2 region of *UmPpc*-1.1:GUS construct causes a complete loss of activity (Addendum figure 1) None of the transgenic plants showed a GUS expression at a histochemical detection level. There could be two explanations for this finding. Either all the four putative CNSs are essential and might work additively for functioning, or the nucleotide distances among the CNSs are vital. To find out the answer further experiments need to be performed in future. However, one thing is absolutely clear from all the experiments done so far is that the CNSs are important elements for activity of *UmPpc* promoter.

Chapter 2

Screening for ethylmethane sulfonate (EMS) mutants of *Arabidopsis thaliana* affected in the anatomy of bundle sheath cells

Part of the research works presented in this chapter has been used for publication in the following manuscript:

"Florian Döring, Kumari Billakurthi, Udo Gowik, Stefanie Sultmanis, Roxana Khoshravesh, Shipan Das Gupta, Tammy Sage and Peter Westhoff (2018). Reporter based genetic screen to identify bundle sheath anatomy mutants in *Arabidopsis thaliana.* Submitted to *The Plant Journal.***"**

Introduction

C_4 photosynthesis is a complex biochemical trait that evolved from their C_3 ancestors by more than 60 times independently in several distantly related groups (Sage, 2016). This complex biochemical reaction is characterized by an efficient CO_2 concentrating mechanism surrounding the ribulose bisphosphate carboxylase/oxygenase (Rubisco) to reduce the adverse effect of photorespiration reaction. The C_4 cycle, in higher plants, relies on the metabolic cooperation of two adjoining chlorenchymatous cell types; the mesophyll (M) and the bundle sheath (BS) cells. The division of labor between the two cell types is supported by a wreath-like structure of M and BS cells around the vascular bundle, the so-called Kranz-anatomy (Haberlandt, 1881). Unlike in C_3 plants, in most of the C_4 plants, CO_2 is primarily assimilated into a C_4 acid by an M cell-specific phospho*enol*pyruvate carboxylase (PEPC). The resulting C_4 acid then shuttles to the BS cells, where it is decarboxylated, and thereby CO_2 is released and concentrated around Rubisco, and subsequently metabolized by the Calvin-Benson cycle (Kanai & Edwards, 1999). In C_4 plants, Rubisco is completely isolated in the BS cells and saturated with CO_2 released by the decarboxylation reaction of the C_4 acids. This CO_2 concentrating mechanism suppresses the photorespiration reaction drastically and produces a favorable condition for Rubisco to operate the Calvin-Benson cycle to its catalytic maximum. With the growing concern about food and energy security, several research groups around the world are working together under an umbrella with an object to install this superior type of photosynthesis trait into existing C_3 crop plants to improve yield (Sage, 2011). In order to achieve this goal, the leaves of C_3 crops must have to go through a lot of anatomical and biochemical modifications.

In C_3 photosynthetic plants, 90 % of the total chloroplasts are located in the M cells within the leaf, therefore, nearly the entire system of photosynthesis takes place in the M cell (Ueno et al., 2006). In contrast, C_4 plants contain an almost equal number of chloroplasts in the BS and M cells. This is because the photosynthesis in C_4 plants is compartmentalized into M cells and BS cells. To facilitate this division of labor, C_4 plants adapted enlarged BS cells with more chloroplasts thereby making the BS cells more pronounced and photosynthetically active (Ueno et al., 2006; Muhaidat et al., 2011; Lundgren et al., 2014). C_4 plants contain not only the increased number of

chloroplasts but also pile other cell organelles like mitochondria and peroxisomes (Brown & Hattersley, 1989) in their BS cells in comparison to the BS cells of C_3 plants. Moreover, in comparison to the C_3 plants, the C_4 plants carry a higher density of plasmodesmata connecting the M and BS cells in order to move the photosynthetic metabolites between the two cell types (Botha, 1992; Danila et al., 2016). All the anatomical changes of the C_4 leaf must have occurred during the evolution of C_4 photosynthesis. It is widely accepted that evolution of C_4 photosynthesis from the C_3 ancestral state has occurred step by step (Sage, 2004; Gowik & Westhoff, 2011; Heckmann et al., 2013). Each small change contributes to the general fitness of the resulting C_3-C_4 intermediate plants. Preconditioning of the proto-kranz anatomy is one of the critical steps in term of an evolutionary trajectory from C_3 to C_4 photosynthesis. This includes an increase in vein density in the leaf, which also results in an increasing number of photosynthetically active BS cells. The increased vein density in the leaf is assumed to be an adaptation to hot and dry climates (Scoffoni et al., 2011; Sage et al., 2012). At this stage of the C_4 evolution, the size of the BS cells also gets enlarged in order to accommodate more cell organelles into it. The well-established proto-kranz anatomy sets a condition for the development of the so-called C_2 photosynthesis. This is achieved by the localization of the glycine decarboxylase complex (GDC) in the BS cells and allowing the decarboxylation of the photorespiratory glycine exclusively in this cell type (Sage et al., 2012). The release of the photorespiratory CO_2 in the BS cells is considered as an evolutionary bridge between C_3 and C_4 photosynthesis. Before the establishment of an active C_2 photosynthesis in the leaf tissue, the BS cells must increase in cell size and chloroplast number. However, the genetic components regulating the development of the BS cell anatomy in C_4 plants remains poorly understood. In order to efficiently introduce the C_4 photosynthesis into C_3 crop plants, detailed knowledge about the genes that are associated with the development of BS cell size and organelle content are required.

Studies with the 5´-flanking sequences of the gene encoding the P subunit of glycine decarboxylase (*GLDPA*) of the Asteracean C_4 species *Flaveria trinervia* showed that the *GLDPA* promoter could keep its BS cell-specificity in the C_3 model species *Arabidopsis thaliana* when fused with a reporter gene (Engelmann et al., 2008; Wiludda et al., 2012). This indicates that the *trans*-regulatory elements responsible for the *GLDPA* BS cell-specific gene expression are already present in the Brassicacean C_3

species *A. thaliana*. In another study, it was found that the BS cell-specific promoter of the *A. thaliana* sulfate transporter gene *SULTR2;2* (Takahashi et al., 2000) maintains its BS specificity in the Asteracean C_4 species *Flaveria bidentis* (Kirschner et al., 2018). This implies that the gene regulatory elements of the BS cells of the C_3 and C_4 dicotyledons plants are at least partly conserved. Therefore, it is assumed that the C_3 model plant *A. thaliana* can be used for discovery of the conserved genes that are involved in the photosynthetic activation of BS cells (Westhoff & Gowik, 2010). In order to discover the genes responsible for BS cell ontogeny and maintenance in C_4 plants, we performed a forward genetic screen in the *A. thaliana.* For this purpose, a genetic background was needed in which BS cells are exclusively marked in the leaf. Therefore, a transgenic *A. thaliana* line expressing Green Fluorescence Protein (GFP) in the chloroplasts of the BS cells was used as a background plant for the experiment (Döring, 2017). An aberrant GFP expression in the primary mutant lines served as a proxy to collect mutants of interest, possibly with altered BS cell anatomy. Mutant lines with the intact reporter gene showing aberrant GFP expression as compared to the non-mutagenized reporter line were collected. Later the microscopic analysis of the selected mutants was carried out to identify if any anatomical changes in the bundle sheath cells and the vascular tissue.

Results

Selection of a BS-labeled reporter line for forward genetic screen

A well-designed forward genetic screen is a powerful tool to study many biological processes and to identify their related genes and regulatory networks. The success of a forward genetic screen depends on an easy, reliable, and a robust primary screening procedure for mutants in which a large number of plants need to be analyzed (Page & Grossniklaus, 2002). By taking the advantages of the forward genetic screen, one of the primary aims of this study was to generate a set of *A. thaliana* mutant lines that were potentially affected in BS cell size or chloroplast numbers. However, the BS cells of C_3 *A. thaliana* are smaller than those of a typical C_4 plant and are only detectable under microscopic analysis. For an efficient and economic primary screen of mutants of interest, an ease BS cell detectable background plant is needed without destructing the leaf tissue. In an established forward genetic screen performed by Döring (2017)

developed a transgenic *A. thaliana* (Ecotype Columbia-0) where the BS cells were labeled with a chloroplast-located GFP protein. The 1571-bp 5´-flanking region of a glycine decarboxylase P protein gene (*GLDPA*) of the C_4 Asteraceaen species *Flaveria trinervia* and the reporter gene *GFP* fused with a chloroplast targeting signal sequence (TP_{RbcS}) were used to develop this BS labeled reporter line (Figure 1A). In *F. trinervia*, *GLDPA* encodes the P subunit of mitochondrial glycine decarboxylase complex, which is a crucial component for photorespiration reaction and exclusively expressed in BS cells (Engelmann et al., 2008; Wiludda et al., 2012). The promoter fragment of this gene is also found to be highly active in the BS cells and vascular tissue of *A. thaliana* (Engelmann et al., 2008; Wiludda et al., 2012). Döring (2017) obtained a homozygous transgenic *A. thaliana* showing a high GFP expression phenotype in the BS cells (Figure 1B). Further longitudinal-section of this transgenic plant revealed that the GFP expression signal could be detected in the chloroplasts of BS cells (Figure 1C). This implies the reporter line could be used as a background plant to generate the easily traceable mutant lines, in which BS cells and its chloroplast are affected.

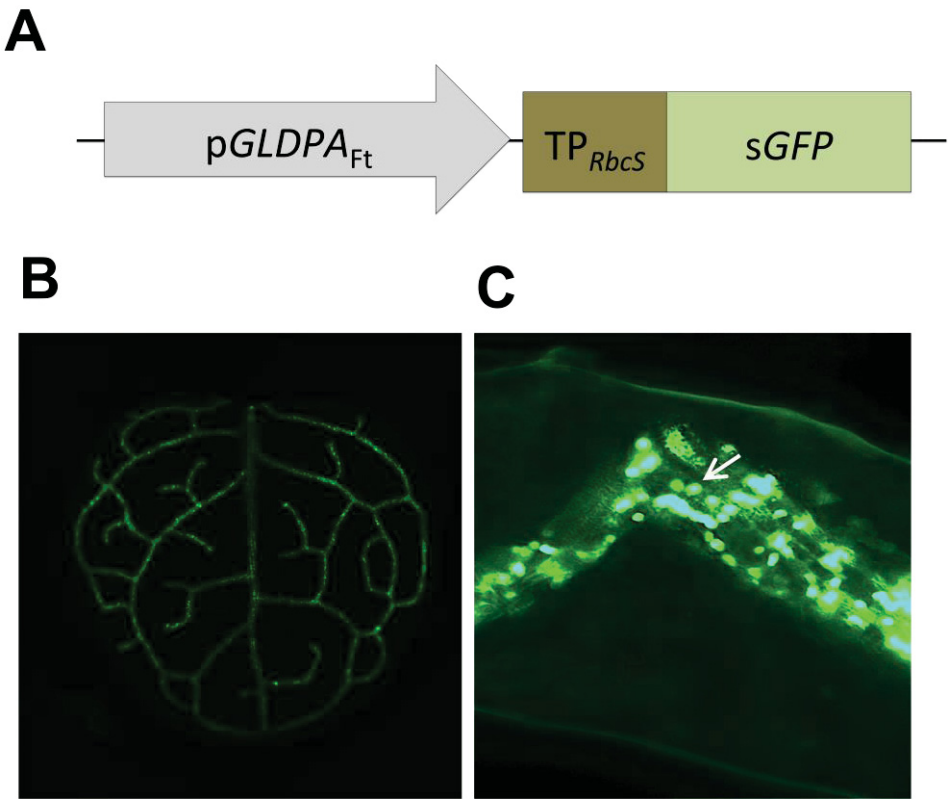


Figure 1: Schematic representation of the BS cell labeled reporter line construction. (A) The construct used to generate the BS cells labeled *GFP* reporter line (Döring, 2017). A chloroplast transit signal (TP_{RbcS}) sequence fused to *GFP* gene

is driven by the promoter of the gene encoding the P subunit of glycine decarboxylase of the C_4 plant *Flaveria trinervia* (p*GLDPA*$_{Ft}$). (B) GFP expression pattern in the BS cells and vasculature of a transgenic plant leaf. (C) Longitudinal section of a vein showing GFP expression in the chloroplast of BS cells and vasculature. White arrow points at a GFP localized single chloroplast. Here, Figure 1C is taken from Döring, 2017.

EMS based forward genetic screen with BS cell labeled background plant

The overall strategy to identify the BS cell related mutants is schematically depicted in the Figure 2. Approximately 40,000 seeds (M1) of BS specific GFP expressing *A. thaliana* were mutagenized by means of chemical mutagen Ethyl methyl sulfonate (EMS) (Kim et al., 2006) and were sown on soil in large flats under greenhouse conditions. It has been observed that 50 % of the total mutagenized M1 seeds did not germinate. The seeds from the survived M1 plants were collected in pools of 25-30 plants. In total, M2 seeds of 300 M1 pools were collected. Next, the seeds of each individual M2 pool were sown on large flats in the greenhouse. It has been observed that approximate 2.2 % of the M2 plants showed a pale chlorophyll phenotype, indicating the EMS treatment was successful (Kim et al., 2006). The whole M2 seedlings were screened for altered reporter gene (*GFP*) expression phenotype in comparison to the GFP expression of the non-mutagenized reporter line. Increased or decreased GFP signal in the BS cells served as the basis for the primary selection of mutants of interest. Since each of the chloroplasts in the bundle sheath cells of reference *A. thaliana* is labeled with GFP, it is assumed that the GFP signal intensity should correlate with the chloroplast numbers or with the bundle sheath size. To this end, we have screened approximately 94,000 M2 mutants and identified a total of 93 mutant lines that showed altered GFP expression in comparison to the GFP expression of the reporter line. All the identified 93 mutants were further analyzed in the M3 generation to confirm the intactness of the aberrant GFP phenotype. In addition to that, GFP expression intensity of the M3 mutants' whole leaves was measured and normalized to the leaf area by employing the software ImageJ (Schneider et al., 2012). M3 mutant lines possessing at least 30 % stronger or weaker GFP signal intensities in comparison to the signal intensity of the reporter lines were selected for further analysis. Thereafter, only 10 out of 94 were found to pass our selection criteria. The T-DNA regions (p*GLDPA*:TP$_{RbcS}$:s*GFP*) of these 10 mutant lines were checked by Sanger di-deoxy

sequencing to be sure that EMS mutagen did not produce any unwanted mutations in the whole reporter gene construct. Indeed, no mutation was identified in the p*GLDPA*:TP$_{RbcS}$:s*GFP* region of our mutant lines (data not shown), indicating aberrant GFP expression was due to the mutation in the genomic region.

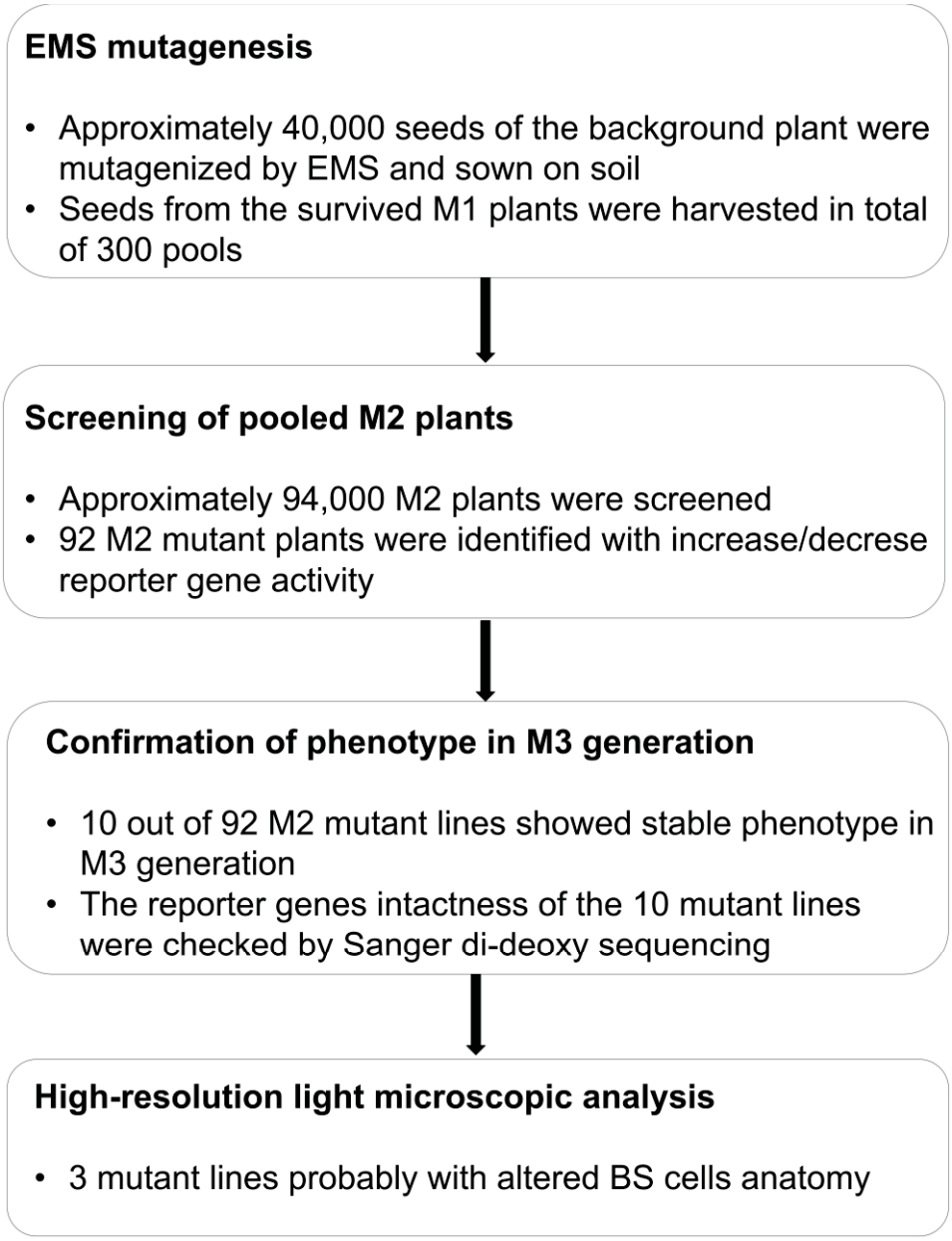


Figure 2: The general workflow of EMS based forward genetic screen with the BS cell labeled *GFP* reporter line.

Analysis of the stable mutant lines showing an aberrant GFP expression phenotype

The EMS based genetic screen resulted in 10 stable mutant lines possessing the aberrant GFP expression phenotype in BS cells. Among these 10 mutant lines, eight mutants showed a stronger GFP signal intensity in comparison to the GFP signal intensity of non-mutagenized reporter plant. These eight mutant lines were designated as 29.1 (+), 42.1(+), 43.1(+), 54.1(+), 62.1(+/Dif), 85.1(+/Dif), 85.2(+) and 88.1(+) (Figure 3). The mutants were named after their discovery from their respective M1 pool number. Within these eight mutants of increased GFP signal, we obtained two mutant lines (62.1(+/Dif) and 85.1(+/Dif)) that additionally possessed a clearly detectable GFP signal in the M cells (Figure 3A). The loss of tissue specificity in these two mutant lines could be due to altered BS cells or M cells development or mutations in genes affecting the transcription and/or posttranscriptional regulation of the p*GLDPAFt* promoter (Engelmann et al., 2008; Wiludda et al., 2012). Among the eight mutants of increased GFP signal, some of the mutant lines showed almost two to three-fold higher GFP signal intensities in comparison to the GFP signal intensity of the non-mutagenized reporter line. For instance, the 54.1(+), 43.1(+) and 85.1 (+/D) lines showed 2, 2.6 and 2.7 folds higher GFP intensity respectively (Figure 3B).

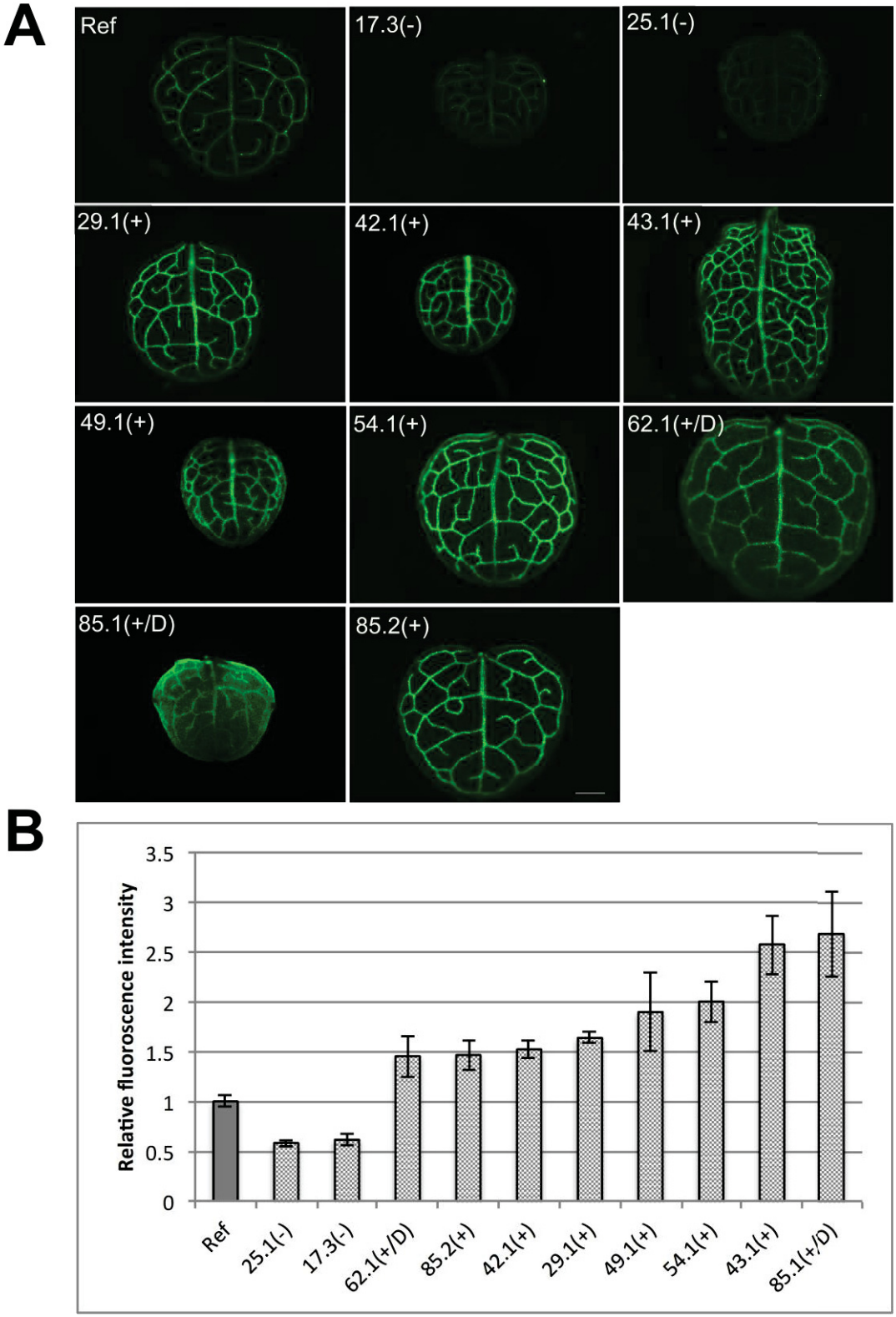


Figure 3: The EMS mutant lines with the altered GFP expression phenotypes. (A) The 10 stable EMS mutant lines with altered GFP expression in comparison to the GFP expression pattern of reference (Ref) line. The fully developed 1st leaf pairs of two

weeks old plants were analyzed for the GFP expression phenotype. Here, the mutant lines with an increased GFP signal intensity are 29.1(+), 42.1(+), 43.1(+), 49.1 (+), 54.1(+), and 85.2(+). The two mutant lines with a diffused GFP signal in the BS cells are 62.1(+/D) and 85.1 (+). A reduce GFP signal intensity was observed in the mutant line 17.3(-) and 25.1(-). Scale bar: 500 µm. **(B)** Relative reporter gene signal intensity of the 10 stable mutant lines. Relative GFP signal intensities was measured with the whole leaves of two weeks old mutant lines and normalized to leaf area. At least 30 plants per mutant lines were analyzed for the measurement.

In addition to these eight mutants of increased GFP signal intensity, we also obtained two mutants (17.3(-) and 24.5(-)) that showed a decreased GFP signal intensity compared to the reference line. Since the GFP of the reporter line is transported to the chloroplast of the BS cells, we assume that any deviation in the GFP signal intensities in the mutant lines is associated with the chloroplast numbers and thus with the bundle sheath size. It could be possible that an increase GFP signal in the mutant lines might be caused by either increase in chloroplasts number in BS cells or an increase in BS cells numbers. In the mutants with decreased GFP signal intensity it could be the case of vice versa. However, at this point, it is not possible for us to clearly address any anatomical alternation of the BS cells based on the changes in the reporter gene expression. In order to correlate the altered GFP gene expression with altered BS cell anatomy, it was important to analyze the mutant lines in more detail with high-resolution microscopic analysis.

Microscopic analysis of the mutant lines and future perspectives

The altered GFP expression phenotype of our mutant lines do not clearly tell us if there are any anatomical alternations in the BS cells. Thereby, we performed a high-resolution light microscopic analysis in our mutant lines to address if any correlation exists between the aberrant GFP signal intensity and BS cell anatomy. At first, we wanted to analyze whether any changes occurred in the BS cells size or number in our mutant lines. To address it, we firstly focused on the high-resolution light microscopic analysis of the leaf cross-sections of our mutant lines. Mutant lines that displayed alternation in the BS cells size or number would be further considered for the ultrastructure analysis by transmission electron microscope (TEM). To carry out the

light microscopic analysis, we chose the six of our best mutant lines based on the strongest phenotypes in terms of GFP signal intensity. Among these subsets of mutants, we included four mutant lines with an increased GFP signal (29.1(+), 42.1(+), 43.1(+) and 54.1(+)), one mutant line with a decreased GFP signal (24.5(-)) and one mutant with a diffused GFP signal (62.1(+/D)). To perform the high-resolution light microscopic analysis, three biological replicates of each mutant line were taken for harvesting the leaf samples. The harvested leaf samples were embedded in resin block and the cross-sections were prepared accordingly (see Material and methods part). In order to compare the cross sections of the mutant lines with the non-mutagenized reporter plant, the leaf cross sections of the reference line were also prepared. For comparison, only the higher order veins (tertiary vein) were considered with respect to the anatomy of the bundle sheath and vasculature. The sample shows a same phenotype in the comparable veins of the three biological replicates were only taken into account for analysis.

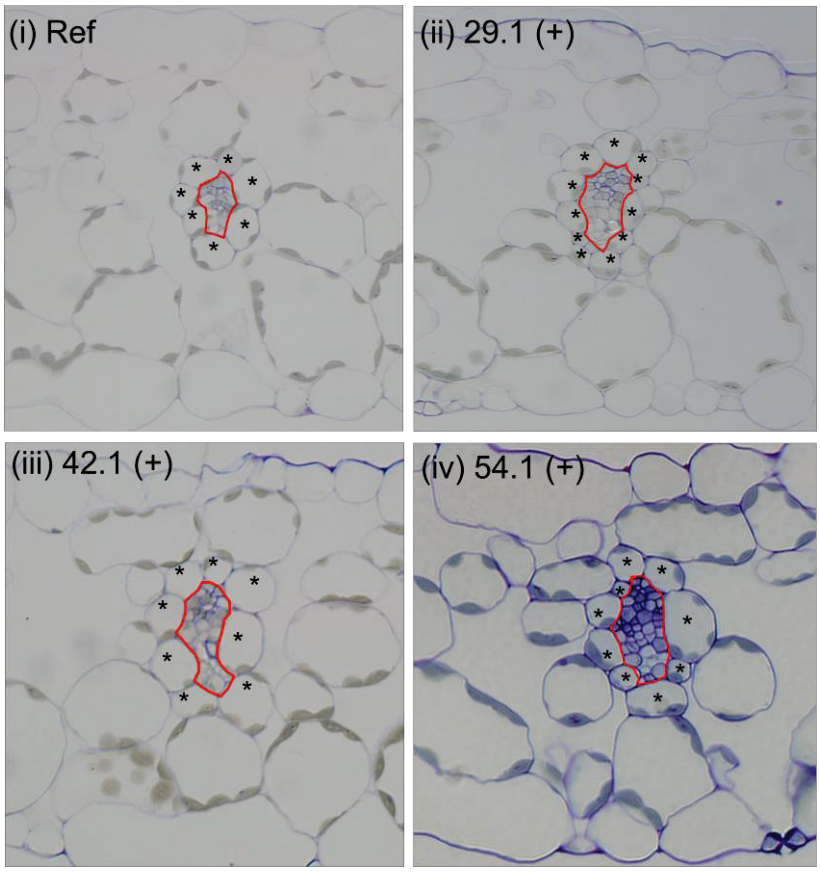


Figure 5: Light micrograph illustrating the resin-embedded leaf cross-sections of the reference and the mutant lines. Here, (i) shows the leaf cross-section of a tertiary

vein of the reference line. (ii), (iii), and (iv) represent the cross-sections of comparable veins of the mutant lines 29.1(+), 42.1(+), and 54.1(+) respectively. The BS cells are marked with asterisks, and the vascular bundle in all sections is framed with a red line.

In the light microscopic analysis, we were not able to generate any conclusion from the mutant lines 24.5(-), 43.1(+), and 62.1(+/D) due to the insufficient comparable veins in their analyzed biological replicates. The cross-sections of at least the two biological replicates of mutant lines 29.1(+), 42.1(+) and 54.1(+) showed an increased number of BS cells as compared to the reference line. However, one biological replicate of each of the three respective mutants did not generate any data due to low sample size and absence of comparable veins. Nevertheless, we obtained three mutant lines from the first round of microscopic analysis that seemed to have more BS cells number in comparison to the reporter line. Additionally, we observed that the increase in BS cells number in these mutant lines is accompanied by an apparent increase in the vasculature per vein. Moreover, the BS cells size in the 29.1 (+) mutant line seemed smaller than the reference line. Since we could not produce a comparable data from all the three biological replicates of each of the mutant lines, it would be too risky to draw a solid statement at this point. Thereby a further confirmation of these mutant lines is warranted in future work.

Discussion and future perspectives

A good understanding of the photosynthetic activation mechanism of BS cells is required for engineering a C_4-like cycle into the existing C_3 crops. A well-designed forward genetic screen could be helpful to unravel the genetic information controlling the photosynthetic activation of BS cells. One of the challenging and laborious tasks in this context is to identify the mutants that are affected in the BS cells and its organelle contents. In this study, we performed a well-established EMS-based genetic screen using *A. thaliana* aiming to identify the mutants of interest (Döring, 2017). Based on the BS cell-specific reporter gene expression intensity as a proxy to screen mutants, we were able to identify 10 mutant lines that are showing a deviated reporter gene expression as compared to the non-mutagenized reference line. In order to correlate the aberrant reporter gene expression with the BS cells anatomy, a high-resolution light

microscopic analysis was also carried out in the six-selected mutant lines. Lack of comparable sample size in the light microscopic analysis did not allow us to address a concrete statement of the result. Nevertheless, three of the mutant lines seemed to have an effect on the BS cells number and vasculature. The observed expansion of the vasculature might be due to an increase in the cell division within the vascular tissue. The ontogenetic relation between the vascular tissue and its surroundings BS cells has been well studied in grass species. It has been found that all the C_3 and many of the C_4 grass species showed a double-sheath surrounding the vascular tissue (Prendergast et al., 1987; Lundgren et al., 2014). The double-sheath is characterized by a layer of non-photosynthetic mestome-sheath encircling the vascular tissue, which is itself surrounded by a layer of BS cells. In contrast to the majority of grass species, some of the C_4 grass, i.e., the NADP-ME type C_4 grass species, do not possess the mestome-sheath, but only a bundle-sheath whose cell walls are equipped with a suberin layer (Brown 1975; Rao & Dixon, 2016). A detailed study performed by Dengler et al., (1985) provided enough evidence about the origin of BS cells in the single-sheath C_4 grass as well as the double-sheath C_3 and C_4 grass species. According to their report, the vascular tissue and its encircling cell-layers of major veins are derived from a procambial initial. Further studies by Bosabalidis et al., (1994) showed that both major and minor veins and its associated BS cells are originated from a single cell lineage in the median layer of leaf primordium. In another study, it was found that the BS cells of the dicots C_3 and C_4 Cleomes originate from more than one layer of ground meristem cells, and only the adaxial BS cells are derived from a procambial initial (Koteyeva et al., 2014). Since the development of BS cells and vasculature in the grass and dicot plants is fully or partially dependent on the same cell lineage, it could be possible that changes in the vasculature are associated with the development of BS cells.

The ultimate goal of this study was to identify the genetic information related to the C_4 BS cells development. Still, there is a long way to go to achieve this goal. Further light microscopic, as well as ultrastructural analysis of our identified mutant lines, is needed to be performed to confirm the obtained phenotype. Once the phenotype is confirmed at the microscopic level, the next step would be to identify the genes related to phenotype. This could be done by a mapping-by-sequencing approach (Schneeberger et al., 2009). A conventional genetic mapping relies on the outcrossing of a mutant line with a diverged genome to produce a segregating mapping population. Since our mutant

screening is depended on the GFP expression in the BS cells, we cannot follow the standard procedure for gene mapping using outcrossing to generate the segregating mapping population. However, it has been reported in many studies that backcross population also produce enough genetic diversity to identify the causative point mutation (Abe et al., 2012; James et al., 2013). Therefore, to generate the segregating mapping population, we have to backcross our mutant lines with the non-mutagenized reference line. The F1 progeny obtained from the backcross population will be self-fertilized, and the resulting F2 population is expected to segregate the recessive mutant phenotype according to the Mendelian law (Sun & Schneeberger, 2015). Finally, the genomic DNA sample from the F2 backcrossed plants could be used for sequencing to identify the causative point mutations. Once the putative candidate genes are identified, the next step will be to verify the genes function with the related phenotypes by performing a reverse genetics study. The whole procedure of the successful screening of BS cell mutants and the identification of putative point mutations is well explained in the Döring et al., (2018, unpublished) and Döring (2017).

Materials and Methods

Plant material

The BS labeled *GFP* reporter line of *Arabidopsis thaliana* (Ecotype Columbia-0) was used in this study as a genetic background for forward genetic screen (Döring, 2017). The Arabidopsis seeds were first surface sterilized for 5 minutes by means of a chloric solution containing 20 % Dan Klorix (Colgate-Palmolive, Hamburg, Germany) and 0.02 % Triton X-100. Then the seeds were washed four times with distilled water. Afterward, the sterilized seeds were incubated at 4 ºC in dark condition for 48 hours. Next, the stratified seeds were sown on soil (Floraton 1, Floragard, Oldenburg, Germany) and the plants were grown under greenhouse condition with 14 hours of light, at a light intensity of ~300 $\mu Em^{-2}s^{-1}$. In addition to the greenhouse condition, the valuable M3 mutant plants were grown in a controlled growth chamber under long day condition with diurnal cycle of 16 hours light and 8 hours dark at a constant temperature of 21-22 ºC. The humidity of the growth chamber was around 40 % and the light intensity 80 $\mu Em^{-2}s^{-1}$.

EMS mutagenesis of the BS labeled *GFP* reporter line and mutant screening

To perform the EMS mutagenesis approximately 40000 seeds (~1.6 g) of BS labeled *GFP* reporter line were taken. The seeds were washed with 0.1 % TWEEN® 20 for 15 min, followed by adding EMS (Sigma-Aldrich, St. Louis) to a final concentration of 0.25 %. Subsequently, the solution was kept on a rotary shaker for 16 hours. Then, the seeds were washed four times with distilled water followed by incubation in a rotary shaker for 1 hour. The seeds were washed one more time with distilled water and subsequently stratified by incubating at 4 ºC in dark condition for 2 days. Next, the seeds were sown on soil and allowed the M_1 plants to grow for maturity under greenhouse condition. M_2 seeds from the M_1 plants were harvested in pools of about 30-50 plants. From each individual pool, approximately 1000 seeds were sown on soil. After two weeks, the M_2 seedlings with first pair of leaf were used for genetic screening. GFP expressions of the M_2 mutant seedlings were compared with the aid of a fluorescence binocular microscope (Nikon SMZ25, Düsseldorf, Germany). Mutants that showed aberrant GFP expressions compared to the GFP expression of the reporter line were selected and grown to next generation. All the primary mutants identified in the M_2 generation were further analyzed in the M_3 generation in order to confirm the individual aberrant GFP expression phenotype.

Isolation and sequence verification of the reporter gene from mutant lines

The genomic DNA was isolated from the M_3 mutant lines according to the protocol of (Edwards et al., 1991). Purified DNA was used to amplify the whole promoter-reporter gene construct (p*GLDPA*:TP$_{Rbc}$S:SGFP) by the Phusion High-Fidelity DNA Polymerase (New England Biolabs). For amplification of whole promoter-reporter gene construct, pGLDPA-Ft-F forward primer and sGFP+SacI reverse primer was used (see Table 1). The agarose gel purified PCR fragments were then cloned into pJet1.2/blunt vector (ThermoFisher Scientific, Waltham, USA) according to the company manual instructions. The positive clones were identified via colony PCR using the pJET1.2 Forward Sequencing Primer and the pJET1.2 Reverse Sequencing Primer of the Kit. Next, the plasmids from the positive colonies were extracted using the QIAprep® Spin Miniprep Kit (Qiagen, Hilden, Germany) and subsequently sequenced by Sanger di-deoxy sequencing (LGC Genomics, Berlin, Germany). List of oligonucleotides used for sequencing of the reporter gene construct are listed in the

Table 1. Mutant lines without any point mutations within the p*GLDPA*:TP*RbcS*:sGFP region were analyzed in this study.

Table 1: The list of oligonucleotides used for amplification and sequencing of p*GLDPA*$_{Ft}$::TP*RbcS*-sGFP. F, Forward primer; R, Reverse primer

Gene specific oligo nucleotides	Sequence (5′ to 3′)	Orientation
pGLDPA-Ft-F	**TACTCCTCTCAACTTTCAA**	F
pGLDPA-Ft-R	**AGTGTAAGATGGGGTCTAA**	R
RbcS.TP+BamHI	**AAGGATCCATGGCTTCCTCTATGCTC**	F
RbcS.TP+EcoRI	**AAGAATTCTTCGGAATCGGTAAGGTC**	R
sGFP+EcoRI	**ATGAATTCATGGTGAGCAAGGGCGAG**	F
sGFP+SacI	**ATGAGCTCTTACTTGTACAGCTCGTC**	R

Light Microscopic imaging

The light microscopic analysis was carried out according to the protocol described in (Khoshravesh et al., 2016). Sample preparation for light microscopic analysis includes series of simple, well-defined procedures. These include (1) tissue fixation and dehydration, (2) infiltration and embedding with resins, and (3) sectioning and staining. The fully developed 2nd leaf pair was harvested from the 4 weeks old mutant plants. Midvein and the edges of the leaf were removed with the help of razor blade and the remaining part of the leaf was cut into 1-2 mm^2 small sections. The small sections were immediately fixed with a fixative solution (1 % glutaraldehyde, 1 % paraformaldehyde, 0.1 M sodium cacodylate) followed by post-fixation with OsO_4. The fixed samples were then dehydrated by a series of ethanol application. After the final steps of dehydration, samples were embedded immediately with the Araldite resin. The resins embedded tissue blocks were then used for cross section with a microtome (Microm HM 330 Microtome) in order to get the sections of 1.5-2 µm thickness. Finally, the cross sections were stained with Toluidine blue and subsequently analyzed on a Zeiss Axiophot microscope equipped with a DP71 Olympus camera and image analysis software (Olympus cellSens, 2009).

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG/GRK1525) through the International Graduate School for Plant Science (iGRAD-Plant) "The dynamic response of plants to a changing environment". I wish to thank Dr. Florian Döring for providing the *GFP* reporter line. I also would like to thank Kumari Billakurthi for making the cross-sections of EMS mutant lines and analyzing them.

Reference

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., ... & Innan, H. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature biotechnology*, ***30***(2), 174.

Bosabalidis, A. M., Evert, R. F., & Russin, W. A. (1994). Ontogeny of the vascular bundles and contiguous tissues in the maize leaf blade. *American Journal of Botany*, ***81***(6), 745-752.

Botha, C.E. (1992). Plasmodesmatal distribution, structure and frequency in relation to assimilation in C3 and C4 grasses in southern Africa. *Planta, **187***, 348–358.

Brown, R. H., & Hattersley, P. W. (1989). Leaf anatomy of C3-C4 species as related to evolution of C4 photosynthesis. *Plant physiology*, ***91***(4), 1543-1550.

Brown, W. V. (1975). Variations in anatomy, associations, and origins of Kranz tissue. *American Journal of Botany*, ***62***(4), 395-402.

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The plant journal*, ***16***(6), 735-743.

Danila, F.R., Quick, W.P., White, R.G., Furbank, R.T., & von Caemmerer, S. (2016). The metabolite pathway between bundle sheath and mesophyll: quantification of plasmodesmata in leaves of C3 and C4 monocots. *The Plant Cell, **28***, 1461–1471

Dengler, N. G., Dengler, R. E., & Hattersley, P. W. (1985). Differing ontogenetic origins of PCR ("Kranz") sheaths in leaf blades of C4 grasses (Poaceae). *American Journal of Botany*, ***72***(2), 284-302.

Döring, F. (2017). *Understanding the Bundle Sheath in C4 Evolution-Forward Genetic*

and Transcriptomic Approaches (Doctoral dissertation, Universitäts-und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf).

Edwards, K., Johnstone, C., & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic acids research*, ***19***(6), 1349.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., ... & Westhoff, P. (2008). The gene for the P-subunit of glycine decarboxylase from the C4 species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis* (C4) and Arabidopsis (C3). *Plant physiology*, ***146***(4), 1773-1785.

Gowik, U., & Westhoff, P. (2011). The path from C3 to C4 photosynthesis. *Plant Physiology*, ***155***(1), 56-63.

Haberlandt, G. (1882). Vergleichende anatomie des assimilatorischen gewebesystems der pflanzen. *Jb. wiss. Bot., **(13)***, 74-188.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P., & Lercher, M. J. (2013). Predicting C4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell*, ***153***(7), 1579-1588.

James, G. V., Patel, V., Nordström, K. J., Klasen, J. R., Salomé, P. A., Weigel, D., & Schneeberger, K. (2013). User guide for mapping-by-sequencing in Arabidopsis. *Genome biology*, ***14***(6), R61.

Jefferson, R. A., Kavanagh, T. A., & Bevan, M. W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO journal*, ***6***(13), 3901-3907.

Kanai, R., & Edwards, G. E. (1999). The biochemistry of C4 photosynthesis. *C4 plant biology*, 49-87.

Khoshravesh, R., Lundsgaard-Nielsen, V., Sultmanis, S., & Sage, T. L. (2017). Light microscopy, transmission electron microscopy, and immunohistochemistry protocols for studying photorespiration. In *Photorespiration* (pp. 243-270). Humana Press, New York, NY.

Kim, S., Lee, D. S., Choi, I. S., Ahn, S. J., Kim, Y. H., & Bae, H. J. (2010). *Arabidopsis thaliana* Rubisco small subunit transit peptide increases the accumulation of *Thermotoga maritima* endoglucanase Cel5A in chloroplasts of transgenic tobacco plants. *Transgenic research*, ***19***(3), 489-497.

Kim, Y., Schumaker, K. S., & Zhu, J. K. (2006). EMS mutagenesis of Arabidopsis. In *Arabidopsis Protocols* (pp. 101-103). Humana Press.

Kirschner, S., Woodfield, H., Prusko, K., Koczor, M., Gowik, U., Hibberd, J. M., & Westhoff, P. (2018). Expression of SULTR2; 2 in the Arabidopsis bundle sheath and vein cells is mediated by a positive regulator. *Journal of experimental botany*.

Koteyeva, N. K., Voznesenskaya, E. V., Cousins, A. B., & Edwards, G. E. (2014). Differentiation of C4 photosynthesis along a leaf developmental gradient in two Cleome species having different forms of Kranz anatomy. *Journal of experimental botany*, ***65***(13), 3525-3541.

Lundgren, M. R., Osborne, C. P., & Christin, P. A. (2014). Deconstructing Kranz anatomy to understand C4 evolution. *Journal of experimental botany*, ***65***(13), 3357-3369.

Muhaidat, R., Sage, T. L., Frohlich, M. W., Dengler, N. G., & Sage, R. F. (2011). Characterization of C3–C4 intermediate species in the genus *Heliotropium* L. (Boraginaceae): anatomy, ultrastructure and enzyme activity. *Plant, Cell & Environment*, ***34***(10), 1723-1736.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., ... & Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of bioscience and bioengineering*, ***104***(1), 34-41.

Page, D. R., & Grossniklaus, U. (2002). The art and design of genetic screens: *Arabidopsis thaliana*. *Nature Reviews Genetics*, 3(2), 124.

Prendergast, H. D. V., Hattersley, P. W., & Stone, N. E. (1987). New structural/biochemical associations in leaf blades of C4 grasses (Poaceae). *Functional Plant Biology*, ***14***(4), 403-420.

Rao, X., & Dixon, R. A. (2016). The differences between NAD-ME and NADP-ME subtypes of C4 photosynthesis: more than decarboxylating enzymes. *Frontiers in plant science*, *7*, 1525.

Sage, R. F. (2004). The evolution of C4 photosynthesis. *New phytologist*, ***161***(2), 341-370.

Sage, R. F., Sage, T. L., & Kocacinar, F. (2012). Photorespiration and the evolution of C4 photosynthesis. *Annual review of plant biology*, ***63***, 19-47.

Sage, R. F. (2016). A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of experimental botany*, ***68***(2), e11-e28.

Sage, R. F., Christin, P. A., & Edwards, E. J. (2011). The C4 plant lineages of planet Earth. *Journal of Experimental Botany*, ***62***(9), 3155-3169.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., ... & Andersen, S. U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature methods*, ***6***(8), 550.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, *9*(7), 671.

Scoffoni, C., Rawls, M., McKown, A., Cochard, H., & Sack, L. (2011). Decline of leaf hydraulic conductance with dehydration: relationship to leaf size and venation architecture. *Plant Physiology*, *156*(2), 832-843.

Sun, H., & Schneeberger, K. (2015). SHOREmap v3. 0: fast and accurate identification of causal mutations from forward genetic screens. In *Plant Functional Genomics* (pp. 381-395). Humana Press, New York, NY.

Takahashi, H., Watanabe-Takahashi, A., Smith, F. W., Blake-Kalff, M., Hawkesford, M. J., & Saito, K. (2000). The roles of three functional sulphate transporters involved in uptake and translocation of sulphate in *Arabidopsis thaliana*. *The Plant Journal*, ***23***(2), 171-182.

Ueno, O., Kawano, Y., Wakayama, M., & Takeda, T. (2006). Leaf vascular systems in C3 and C4 grasses: a two-dimensional analysis. *Annals of Botany*, ***97*(**4), 611-621.

Westhoff, P., & Gowik, U. (2010). Evolution of C4 photosynthesis—looking for the master switch. *Plant Physiology*, ***154***(2), 598-601.

Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., ... & Westhoff, P. (2012). Regulation of the photorespiratory *GLDPA* gene in C4 *Flaveria*: an intricate interplay of transcriptional and posttranscriptional processes. *The Plant Cell*, ***24***(1), 137-151.

Acknowledgements

I wish to thank, first and foremost, **Prof. Dr. Peter Westhoff** for giving me an opportunity to work in his research group, for his supervision, valuable advice and support.

I would like to thank **Prof. Dr. Maria von Kroff Schmising** for being member of my Ph.D. committee and for her valuable review of this dissertation.

I cannot find words to express my gratitude to **Dr. Udo Gowik**, **Jan Emmerling**, **Dr. Florian Döring**, **Kumari Billakurthi,** and **Dr. Stefanie Schulze** for their valuable advices and especially for their incredible patience. The way they see and interact with the science, helped me to stay motivated in this field of knowledge.

Dr. Sandra Kirshner, **Zahida Bano** and **Jasmin Mehsing** for your kind support in many of my personal German bureaucratic issues.

I would also like to thank **Monika Streubel**, **Susanne Paradies**, and **Maria Koczor** for their help and valuable advices during my lab works.

I would like to thank all the **iGRAD-Plant fellows** for their valuable advices and help during the iGRAD-Plant mini seminars. I would not forget those fun moments that I had spent with you guys in the iGRAD-Plant retreat program, Christmas party, and so on.

I would like to offer a heart-felt thank you to iGRAD-Plant course coordinator **Dr. Petra Fackendahl** for helping me with all the complicated paper-works and making my life easier in Germany.

I owe my deepest gratitude **to Prof. Federica Brandizzi** for accepting me as a short time research scholar at Michigan State University and her research group for making my research stint memorable.

I wish to thank my beloved mother **Minati Das Gupta**, family members, and all of my friends for their inspiration and supports.

Finally, I could not have made it this far without support and love from my beloved wife **Prapa Dhar** and my son **Sreontho Das Gupta.** Studying in Germany would have not been possible for me without their encouragement.